



The geography of connectivity: a review of mobile positioning data for economic geography

Andreas Erlström¹ · Markus Grillitsch^{1,2,3} · Ola Hall¹

Received: 17 February 2021 / Accepted: 5 June 2022 / Published online: 12 July 2022
© The Author(s) 2022

Abstract

Connectivity between and within places is one of the cornerstones of geography. However, the data and methodologies used to capture connectivity are limited due to the difficulty in gathering and analysing detailed observations in time and space. Mobile phone data potentially offer a rich and unprecedented source of data, which is exhaustive in time and space following movements and communication activities of individuals. This approach to study the connectivity patterns of societies is still rather unexplored in economic geography. However, a substantial body of work in related fields provides methodological and theoretical foundations, which warrant an in-depth review to make it applicable in economic geography. This paper reviews and discusses the state-of-the-art in the analysis of mobile phone and positioning data, with a focus on call detail records. It identifies methodological challenges, elaborates on key findings for geography, and provides an outline for future research on the geography of connectivity.

Keywords Connectivity · Mobile phone data · Spatial mobility · Social networks · Regional development · Mobile positioning data

JEL Classification C00 · C18 · O18 · R00 · R14

1 Introduction

Connectivity between and within places is one of the fundamental cornerstones of economic geography (see e.g. Farber et al. 2013; Storper and Venables 2004). Rather than co-location, it is connectivity between people that promotes interactions

✉ Andreas Erlström
andreas.erlstrom@keg.lu.se

¹ Department of Human Geography, Lund University, Lund, Sweden

² CIRCLE – Centre for Innovation Research, Lund University, Lund, Sweden

³ Inland Norway University of Applied Sciences, Lillehammer, Norway

and learning in urban spaces, between urban spaces, but also between urban and more remote locations (Farole et al. 2011; Grillitsch and Nilsson 2015; Martin et al. 2018). However, connectivity is not disconnected from co-location. A majority of our social interactions occur locally in spatially embedded networks (Bettencourt et al. 2010), where research accentuates the effects of ‘local buzz’ for firms performance (Bathelt et al. 2004) and the scaling effects of local density for learning and development of human capital (de la Roca and Puga 2017). Connectivity is therefore central to the geography of the economy (Glaeser and Maré 2015) and relates both to organizations and individuals.

However, the data and methodologies used to capture connectivity have been limited due to the difficulty in gathering and analysing detailed observations in time and space about how people interact and move across space, e.g. relying on time geography diaries that are unfeasible on large scales (Farber et al. 2013; Scholten et al. 2012). Mobile phone data potentially offer a rich and unprecedented source of data, which is exhaustive in time and space and closely follows movements and partly communication activities of individuals (Ahas et al. 2007a, b; Kwan 2007; Steenbruggen et al. 2015). However, to the best of our knowledge, research using mobile phone data to measure connectivity in the field of economic geography has been limited, while it has been a more prominent feature in contemporary research among related fields. This paper, therefore, provides a methodological overview of how mobile phone data has been used in studies related to economic geography, elaborates on key findings for economic geography, identifies the methodological challenges, and outlines opportunities for future research on the geography of connectivity using mobile phone data.

Since the late twentieth century and the introduction of the computer chip, the world has increasingly become digital. In most of everyday life, individuals interact with devices that leave digital footprints, but most importantly allow for communication across space, which is unprecedented in human history. Social networks are no longer mainly defined within spatial scales, where interactions between individuals would occur on the streets and corners (Jacobs 1969). Communication technology cuts across scales and allows for networks and structures to transcend distance with greater impact than before. This increase in the opportunities for communications across scale have altered the geography of societies, allowing self-organizations that no longer function solely through physical connectivity but through increasingly ‘aspatial’ connectivity.

Mobile phones connected to cellular networks have been essential in this development. From the start, mobile phones through their connection to the GSM network have provided individuals with the ability to stay connected to their social networks regardless of their geographical proximity, allowing both the maintaining and development of social networks across scale. Since the turn of the millennia the use of mobile phones for communication has grown exponentially, which has resulted in a society wide diffusion of the smartphone that allowed for far more applications than mainly communication (Steenbruggen et al. 2015). The development of applications such as WhatsApp, Google maps, Instagram, with increasingly location-based services, from providers such as Twitter, Facebook, Google has increased the utility of the mobile phone and consequently the everyday use. Subsequently, the amount

of data generated, as well as the variety of methods that can be employed to collect data, have increased.

Not surprisingly, the increasing data generated from people's everyday usage of the phone, whether being passive (GPS tracking, WiFi probing, ID tracing) or active (geolocating, calls or texts), have seen increasing interest from researchers. Topics have ranged from mobility analysis (see e.g. Cottineau and Vanhoof 2019; Pappalardo and Simini 2018; Vanhoof et al. 2018), mobile social network analysis (e.g. Gaito et al. 2017; Mamei et al. 2018), tourism (e.g. Ahas et al. 2007a, b; Raun et al. 2016), social influence (e.g. Peng et al. 2017a, b; Singh and Ghosh 2017), literacy (e.g. Blumenstock et al. 2015; Schmid et al. 2017), development, or segregation and poverty (e.g., Hedman et al. 2021; Hernandez et al. 2017; Östh et al. 2018). Furthermore, some studies have developed methods for analysing mobile phone data together with socio-economic factors, using aggregated data. The studies are evidently growing in number and variety, and whilst there are some reviews, they either focus on specific data sources, such as GPS data (e.g. Wu et al. 2016), or it has focus on specific outcomes, such as social networking big data (Peng et al. 2017a, b). Others are dated and are mainly encompassing the early stages (Raento et al. 2009; Ratti et al. 2006). Existing reviews examine the potential of big data for social sciences in general (Lazer and Radford 2017) and its benefits to urban managements and policies (Steenbruggen et al. 2015). However, a state-of-the-art review and discussion about the potential contribution of mobile phone data to study the geography of connectivity specifically is missing.

This paper provides a comprehensive analysis and discussion of methodologies, limitations, and opportunities of mobile phone data for economic geography. The focus lies on examining how big data in the form of mobile phone data could improve the empirical foundations for the analysis of connectivity within the field of economic geography. To this end, we conducted a systematic literature review applying the methodology developed by Tranfield et al. (2003). In total, we screened 427 articles obtained from Ebscohost and 189 articles from Web of Science. The detailed review included 140 articles, which we evaluated and summarized on four dimensions: (a) research questions and the literature the article contributes to, (b) methods of combining spatial mobile phone data with data relevant for economic geography, (c) key findings, and (d) limitations. Section 2 describes how the literature review was conducted. Section 3 discusses methodological approaches for analysing mobile phone data and linking it to other data sources, as well as limitations. Section 4 synthesizes the main findings about human mobility, social networks, and aggregate patterns of human mobility and social networks. Section 5 elaborates on the prospects of using mobile phone data to capture connectivity and learn about its causes and effects in the field of economic geography.

2 Background: mobile phones and connectivity

Mobile phones are a fundamental pillar in communication technologies that have changed our everyday life. Activities such as making a call, reading e-mails, accessing social media or websites, have become so closely linked with the mobile phone

that it has come to reflect much of our social connections. Moreover, mobile phone data leaves a “digital trail” that follows individuals’ movements and their communication activities. This digital trail offers several ways for researchers to study spatial mobility and communication at both individual and aggregate scales and across different time-periods (Raento et al. 2009). In short, mobile phone data can be summarised into three types of data: (1) mobile location data (such as GPS data, call detail records, or tracing data), (2) communication data (online or offline communication between cellular phones), and (3) application data. These in turn can be passively or actively gathered by researchers through various methods (Ahas et al. 2007a, b). This paper will focus on the first two types of data, with a focus on location data as a remote sensing of human activity or communication space. This means that the dataset does not require direct or active intervention from the researcher in order to be generated. Nevertheless, some methods also actively generate sensing data by tracing a sample of mobile phones using the operators register (e.g., Ahas et al. 2007a, b; Ahas et al. 2010; Raun et al. 2016).

Mobile location data records the position of mobile phones in time and space. One approach to gather such data is to utilize the GPS-system built in today’s smartphones (Raento et al. 2009; Yadav et al. 2014). A more uncommon approach is identifying physical proximity between phones through Bluetooth scans. This allows recording the encounters that a subject has over the duration of the study (Eagle et al. 2008; Raento et al. 2009). However, the most common approach to gather large sensing data is through base receiver stations (BTS), also termed cell-towers, to generate location data of mobile phones (Frias-Martinez et al. 2012). Mobile phone operators register when a phone contacts other devices (through calls or texts) or when it uses internet connection (Lu et al. 2017; Rodriguez-Carrion et al. 2018). Furthermore, the BTSs can also be used as a mean of actively tracing the mobile phones’ locations. Phone operators continuously register and store the ID-number of smartphones connected to their customers, and this ID-number can be used to ping them over the communication network (Ahas et al. 2007a, b).

Empirical studies frequently use call logs and location data, which often become available in Call Details Records (CDR) (Lazer and Radford 2017; Teng and Chou 2007). This metadata is typically generated when a phone contacts a cell tower for the purpose of transferring a call or a text (Pappalardo et al. 2015). The advantages of CDR primarily relate to the collection process. Call logs and location data can be accessed indirectly without the need of user interaction beyond the usual use of mobile phones. CDR therefore avoid biases related to individuals’ perceptions. Perceptions play a role, for instance, when users are asked to log their social network. Individuals may record the networks that are perceived important instead of recording actual behaviours (Eagle et al. 2008). CDR are often geographically processed through their connection with the BTSs, which are geo-coded and cover a certain geographical area (Hernandez et al. 2017; Moyano et al. 2012; Pappalardo et al. 2015; Vanhoof et al. 2018). An alternative to the classical CDR is mobile phone location data. In contrast to CDR such data is not only generated when calls or texts are conducted but also whenever the phone uses BTSs for access to e.g., internet. This produces a more detailed dataset, which is less influenced by individual agency (Lu et al. 2017; Rodriguez-Carrion et al. 2018).

Communication data is generated when a user utilizes the phone to contact other users. This involves metadata on when (timestamp), where (through location data), who (through mobile phone ID) and how long the contact (duration of the connection) lasted as well as more detailed accounts about the contents, recording of the call or the content of the text (Eagle et al. 2008, 2009; Raento et al. 2009). As such it bears resemblance to the types of data that is used for mobile location data, such as CDR. However, whilst the former is mainly interested in the phone's interaction with the BTS, the latter is mostly focusing on the connection to another device and does not necessarily follow its location in space (see e.g., Peng et al. 2017a, b). Communication data can be of varying quality. Some datasets make only use of the volume of calls between phones, where phones are treated as nodes and volume of calls as weighted bi-directed edges. Others are collecting the content of the communication. Both extremes provide with different challenges and limitations in the analysis, data management, and research ethics.

The generation of mobile phone data exploded with the transition to smart-phones. The first widely used mobile phones mainly provided with the functionalities of making calls or using the short message services (SMS) through the GSM network. The first mobile phones diffused rather slowly, were taken up mainly by the younger population, and the applications were limited. Since then, incremental and radical innovations have improved the mobile phone as a tool in everyday life; most notably it has increased the variety of applications, with increasing possibilities for communication through different medias, location-sharing through the use of inbuilt GPS-systems and compass, sensing data such as built-in functions to check on biometrics (Ahas et al. 2007a, b). With the shift in use of mobile phones from mainly making calls or sending texts into a device that has become an essential part of everyday life, where individuals of most ages mostly always carry it with them, have consequently caused an increased generation of data (Deville et al. 2014). The everyday use made the positioning data more accurate than it was in the beginning (Steenbruggen et al. 2015). As such the potential benefits for research increased over the last decade and exploring its potential for research in economic geography has therefore become more relevant.

3 Methodology for a systematic literature review

The literature review strictly follows transparent processes by systematically going through a cyclic process of obtaining, screening, and evaluating articles—in line with the methodology suggested by Tranfield et al. (2003). Since the review deals with an emerging and interdisciplinary field, we combined multiple criteria in the selection process (Step 1) to cover exhaustively relevant articles. First, the articles need to combine spatial mobile phone data with keywords of relevance for economic geography as shown in Table 1. The keywords were combined using a Boolean search process that used the command “AND” for the selected keywords, together with “NOT” for the omitted keywords. The first column showcases the keywords used to limit the search to studies of mobile phone-generated data, this approach has been labelled in a variety of ways, but these are

Table 1 Table of Keywords used to collect relevant articles

Mobile phone keyword	Selected keywords		Omitted keywords ("NOT")
"Call Detail Record*"	"Socio-economic"	OR "Development"	"Climate Change"
OR "CDR*"	OR "Socio economic"	OR "Innovation"	OR "nature"
OR "Mobile Phone Data"	OR "Socioeconomic"	OR "periphery"	OR "Healthcare"
OR "Mobile Phone Records"	OR "Income"	OR "rural"	OR "Epidemiology"
OR "Phone Data"	OR "Employment"	OR "urban"	OR "Anthropology"
OR "Mobile Positioning Data"	OR "Unemployment"	OR "Deprivation"	OR "Waste"
	OR "GDP"	OR "Economic characteristics"	OR "Psychology"
OR "Smartphone"	OR "Well-being"	OR "Human Mobility"	OR "Imagery"
OR "Location-based Servic*"	OR "Wealth"	OR "Commuting"	OR "Microelectronics"
	OR "Poverty"	OR "Social Capital"	OR "Transport* simulation"
	OR "Education"	OR "Social Network"	OR "Transport Mode"
			OR "Traffic Flow"
			OR "Tourism"
			OR "Parks"
			OR "disaster*"
			OR "Care"

some of the most frequently occurring in the literature. The second column aims to limit the search to articles dealing with economic geography and especially regional characteristics. Therefore, the keywords chosen tried to fit studies that either deal with the question of connectivity, the question of urban–rural differences, or the question of socio-economic characteristics (linking to the growing body of literature on the geography of inequality). Finally, the study also made use of the NOT function to omit research that incorporated mobile phones but did so with a focus irrelevant to our aim. Whilst the NOT function should be used with care as to not unconsciously exclude papers by using too general terms, the study found it to be called for as mobile phone data has been excessively used in other fields and consequently the array of literature produced, if the NOT-keywords would be omitted, would be too great to fully encompass.

Second, the review uses a defined time frame, articles published between 2008 and 2019.¹ This is due to mobile phone data being inherently linked to the emergence of mobile phone technology and in particular the following two factors: First, for mobile phone data to give a representative picture of a population's mobility or its calling patterns, and the correlation with socio-economic factors, it needs to have good coverage of the studied population at an aggregated level. Such a dispersion of mobile-phone technology varied between regions. The most advanced countries in this respect had a good coverage from the start of this millennium (Lazer and Radford 2017; Raento et al. 2009). Second, the shift in use of mobile phones from mainly making calls into a tool that has become an essential part of everyday life, where individuals of most ages often carry it with them, did not start until the beginning of the 2000s. Predominantly, it started with the rise of the smartphone and later on became more widespread with the release of the first iPhone and Android phones. The importance of this second factor is that this shift also changes how mobile phone data mimics an individual's mobility and communicative pattern (Raento et al. 2009). After the transition, mobile phone data follows more closely individuals' everyday life, rather than particular points in time. Consequently, the time frame was set to be articles with publication date after 2004, which cover empirical material from the early 2000s.

Third, the review aims at including research that critically examined the methodology of analysing mobile phone data, its strength and weaknesses, biases and challenges. These do not necessarily overlap with research that fits within the search criterion in Table 1. Therefore, we also tracked references that provided the methodological basis for the respective paper and included it in the list of articles. This process was repeated until the methodological origin was found and we exhaustively covered all relevant articles. However, the breadth of methodological applications and traditions found led us to the decision to provide a more general overview of methods relevant for the use of mobile phone data in the field of economic geography in line with the main aim of the paper.

In Step 2, we screened the articles by reading the abstracts and the methodology section in order to establish whether spatial mobile phone data was analysed in combination with data pertinent to the keywords identified in Table 1. Studies of mobile phone data have contributed to three main themes relevant to the aim of this study: human mobility, social networks, and aggregate patterns of human mobility and social networks. Human mobility is a crosscutting theme, which often co-occurred

¹ Ebscohost and Web of Science, covering partly overlapping and partly complementary academic databases, were used. The search applied the same search procedure with some exceptions. Ebscohost allows for the Boolean search to be conducted on the entirety of the texts, whilst Web of Science only allowed for a search in specific segments of the articles, including keywords, title and abstract. Ebscohost and Web of Science further gave opportunities to limit the search to specific databases or subject areas. The selections made depended on the options available. Web of Science allowed for a delimitation based on subject areas whereas Ebscohost allowed to select among a variety of databases. The search process with Ebscohost resulted in 427 articles whilst Web of Science provided 189 articles. The fact that Web of Science generated fewer results than Ebscohost was to be expected since it only spanned one database, only allowed search in specific sections of the article, and had better options for limitations.

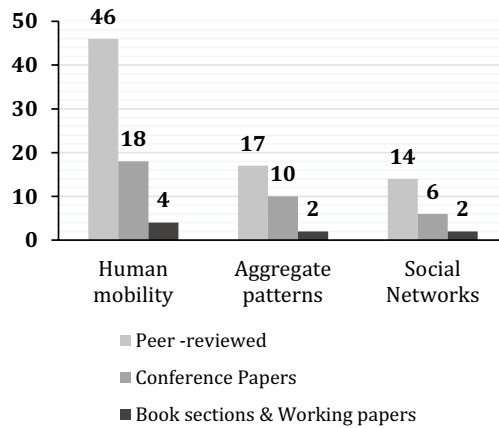


Fig. 1 Papers by topic and type of source

with the other themes. Therefore, articles were only classified under this category if their main purpose was to investigate human mobility with mobile phone data, in total 68 articles. Twenty-two studies focused on spatial differences of social networks. Twenty-eight articles investigated aggregate patterns of human mobility and social networks (Fig. 1).

In Step 3, we evaluated the methods, the data, and the nature of the findings of each article. If these were not adequately presented, the article was classified as inadequate and discarded. Adequacy of presented data and methods requires transparency of the steps in handling and analysing data, as well as a presentation of the nature of the data. In the end, 122 articles were included in this review. The evaluation of these articles zoomed in on the following four dimensions: (a) research questions and the literature the article contributes to, (b) methods of combining spatial mobile phone data with data relevant to economic geography, (c) key findings, and (d) limitations.

4 How is mobile phone data used?

Mobile Phone Data or sensing data is used at the level of individuals and/or in aggregated form (Raento et al. 2009). Aggregated data implies that the location data of individual users have been aggregated to either the Voronoi tessellation or some other form of geographical unit for the purpose of facilitating analysis of e.g. socioeconomic status (Xu et al. 2018), or spatial communities (Gao et al. 2013) at those particular scales. When using aggregated data, each unit is treated as a node in wider network of interactions, measured through mobile phone data. Whilst this study focuses on the literature using aggregated sensing data, the individual data that some

studies made use of holds relevancy to the field of economic geography, especially for studies that focus on case-studies of individual sectors, firms, or clusters. These two forms of sensing data can then be used in different ways to link with other datasets. What follows will be a brief outline of these two approaches before we discuss how the literature have used mobile phone data together with other data, such as socio-economic data. This we then expand on by discussing the way human mobility and connectivity have been operationalized in research using mobile phone data.

Studies on the *individual level* use mobile location data for three major applications that relate to connectivity. Firstly, individual data is used to study the activity space within urban environments following individual's movement in space and time (e.g., Jiang et al. 2012; Scherrer et al. 2018; Yuan and Raubal 2016). This is for example prevalent in tourism studies (see Shoval and Ahas 2016). Other studies analyse activity patterns of humans with the aim to deduce home-work patterns (Ahas et al. 2010), fit algorithms that predict movement patterns (Dashdorj et al. 2018; Doyle et al. 2019; Hoteit et al. 2014), or aim to increase the accuracy of the sensing data (Chen et al. 2018; Rodriguez-Carrion et al. 2018; Zhou and Huang 2016). Secondly, individual data is used to infer social networks from the sensing data (Peng et al. 2017a, b). Here phones are treated as nodes in a network and whilst some have a spatial perspective on the social networks (e.g., Doyle et al. 2019; Onnela et al. 2011; Phithakkitnukoon and Smoreda 2016; Puura et al. 2018), it is not uncommon that these studies leave out spatial patterns in their analysis (e.g., Peng et al. 2017a, b; Werayawarangura et al. 2016). Thirdly, some studies combine sensing data with other datasets such as physical proximity to others or phone surveys with individuals in the dataset (Eagle et al. 2008). In that way, observations about individuals' interactions with their mobile phone is merged with socio-economic information provided by the individuals through surveys or other data sources (see e.g., Engelmann et al. 2018; Fixman et al. 2016; Jahani et al. 2017). The last approach often incorporates the same aims as the earlier two.

The studies using *aggregated data* vary in their application. Especially since the aggregation to Voronoi polygons allows the incorporation of a variety secondary datasets, such as socio-economic data, that are unrelated to the collection of the mobile phone data. In order to combine mobile data with socio-economic data available for administrative territories, the mobile data is further aggregated from Voronoi tessellations to fit the area of the socio-economic data (Cottineau and Vanhoof 2019; Frias-Martinez et al. 2012; Pappalardo et al. 2015), showing the versatility of mobile phone data. Using these Voronoi tessellations polygons, location data is geocoded into regions in such a way that researchers are able to analyse and compare urban and regional patterns (Chi et al. 2016; Šćepanović et al. 2015; Vanhoof et al. 2018; Wang and Kilmartin 2014; Yuan and Raubal 2016). Voronoi polygons are created so the generating points, the masts, are closer to their polygon defining points than any other polygon points, thus creating a plane of areas defined by distances.

Whilst there are a variety of potential applications using aggregated data the ones that focus on physical or social connectivity are primarily characterized by two approaches. Similar to individual studies, certain applications focus on the use of urban space (or differences of activity space across regions). These studies do

not necessarily make use of different datasets, but instead focus their approaches on the use of aggregated mobile positioning data. For instance, Eagle et al. (2009) and Chi et al. (2016) found urban–rural differences in mobility and communication patterns when comparing aggregated CDR data. Ahas et al. (2015) as well as Amini et al. (2014) studied the differences in the spatial patterns observed from mobile phone data across countries, and Chi et al. (2016) as well as Šćepanović et al. (2015) showed how human mobility and communication relate to the local context. The other approach is the application of mobile phone data together with other datasets (Castillo et al. 2018; Cottineau and Vanhoof 2019; Mao et al. 2015; Schmid et al. 2017). For instance, Eagle et al. (2010) focused on the link between network diversity and economic development looking at aggregated communications patterns in Great Britain, and Östh et al. (2018) studied how patterns of urban inequality in Stockholm changed when incorporating the daily rhythm patterns of the population. For most of these studies, the mobile positioning data used has been CDR and several of these studies have potential relevance to the question of connectivity where they not only used mobile phone data alone but together with other datasets.

The process of merging mobile phone data with *other datasets* differs between studies using individual and those using aggregated data. Several methods have been used to link individual CDRs with supplementary data on individual level (see e.g., Blumenstock 2018; Eagle et al. 2008; Järv et al. 2015). This has been done by collecting personal information with surveys, mobile phone applications or by accessing more detailed contract information from the telecommunication provider. From contract information Jahani et al. (2017) and Järv et al. (2015) collected demographic information such as gender, age and preferred language, which Järv et al. (2015) used as a proxy for ethnic groups aimed at exploring segregation in Estonia from differences in activity spaces. For economic data, Engelmann et al. (2018) used m-money transactions provided by the telecommunication provider in Tanzania that included the user ID, timestamp and transaction amount, etc. to infer socio-economic status of individuals. This, they argued, would outperform CDR in e.g., predicting socio-economic status. Fixman et al. (2016) used bank information for a subset of users in the CDR dataset to extract income levels. The use of phone surveys was also used to infer economic data for a subset of the population in Afghanistan and Rwanda studied by Blumenstock (2018). In the end, these processes of merging supplementary data with CDR on individuals require that both datasets can be joined, typically through the telephone ID.

On an aggregated level, various types of data have been linked with mobile phone data. Common datasets include demographic or socio-economic data collected through national household surveys or individual censuses. The common attribute is that in order to link CDR with demographic or socio-economic data on areas, the CDR needs to be aggregated to the scale the socio-economic data was collected on (Castillo et al. 2018; Cottineau and Vanhoof 2019; Pappalardo et al. 2015). For instance, Frias-Martinez et al. (2012) used socio-economic data for districts within a city in an emerging economy in Latin America during 2010 to measure the level of socio-economic development. Cottineau and Vanhoof (2019) used census data for France to create a variety of delineations of urban environments in order to relate

CDR and socio-economic status to the organization of cities. However, you can also link individual phone users to areas rather than aggregating, in that case research is dealing with multilevel datasets. An example is a study of Singapore where Xu et al. (2021) joined together mobile phone data, income levels and housing prices to uncover spatially embedded social networks. They did so by connecting users to a place of residence, inferring a socio-economic status, and could from there link its connectivity to other areas by calling patterns. Similarly, a comparison between Boston and Singapore saw socio-economic status be linked with individual phone user through similar method of inferring place of residence in the mobile phone data. In this case, it was used to study the relationship between human mobility and socio-economic status (Xu et al. 2018).

The majority of the studies in the review made use of CDR to study either communication networks or human mobility. Some made use of other location-based data generated by mobile phone, such as mobile phone tracing. However, with clear majority the data used was generated by some form of call detail record. The popularity of using CDRs to capture human mobility or connectivity can to some extent be attributed to the methods of collecting and processing CDR data. Since CDRs are connected to cell towers, which are geo-coded and represent a defined geographic area, CDRs are thought to reflect the movement patterns of users making CDRs well-suited for analysing human mobility (Lazer and Radford 2017; Mota et al. 2015). Only four variables are needed to analyse human mobility with CDRs: the user ID, the base transceiver station (BTS) ID, the coordinates of the BTS, and the timestamp of the interaction. Furthermore, it has been argued that previous methods of gaining knowledge about the flow of people in an urban environment such as public transportation surveys have some noticeable flaws that mobile phone data could work around (Calabrese et al. 2013; Kung et al. 2014).

Methodologically, human mobility patterns are analysed with CDR using indices based on three types of measures: (1) the travel distance, (2) the range of individuals' activity space, and (3) the heterogeneity of travels (Lu et al. 2017). Travel distance, also sometimes used for mobility volume, often works with the total Euclidean travel distance of users and is the most basic of mobility indicators (Lu et al. 2017). The range of activity space captures the area in which individuals move, and according to Yuan and Raubal (2016, p. 1604) aim to reflect both 'external descriptive statistics (e.g., shape, size) and internal structures (e.g., regularity)'. There are a number of methods that are employed for measuring activity space. One that is recurrently used is the radius of gyration, which also has been used as an index for mobility volume, that measures the spatial spread of locations visited (González et al. 2008; Hoteit et al. 2014; Lu et al. 2017; Pappalardo et al. 2015). Other methods of measuring activity space are standard deviation ellipses, minimum convex hull geometries, or daily potential path area (see Kwan 1999; Sherman et al. 2005; Yang et al. 2016). Heterogeneity of travels is far less defined but often employs mobility entropies, which will give insights to the internal structure of individuals' activity space (Lu et al. 2017; Yuan and Raubal 2016). The foundation lies in modelling the diversity of locations visited. The mobility entropy will be high when individuals conduct many different trips with changing origins and destinations and low when an individual mainly goes through small set of recurring trips (Pappalardo et al. 2015).

On the basis of these indices a variety of methods have been employed to study the similarities of mobility patterns. These range from relatively simple ranking of indices (Becker et al. 2011), to more complex methods that use profiling algorithms (Thuillier et al. 2018), or spatio-temporal edit distance algorithms (Yuan and Raubal 2014). This has then been used to both understand and categorize space by the activity patterns of individuals (see e.g., Ahas et al. 2015; Dash et al. 2015; Manfredini et al. 2013) and to classify and group communities (Becker et al. 2011; Thuillier et al. 2018). The aim is then to not only identify the different home-locations and workplaces in commuting patterns but also to identify differences in staying locations in the urban environment as well as the constraints of activity spaces (Dashdorj et al. 2014, 2018; Hoteit et al. 2014; Järv et al. 2015; Yang et al. 2016).

Beyond human mobility, the communication details from CDR have been used for social network analysis (Eagle et al. 2008; Raento et al. 2009). This set of research started in connection with traditional social network methods that used self-reporting surveys in order to map social networks or social capital (Eagle et al. 2008; Ghosh and Singh 2018). CDR provide researchers with the ability to infer the ties between nodes, as well as the edges and the links, by using e.g. number of calls between users (Calabrese et al. 2011) or by using the duration of the calls between contacts (Onnela et al. 2011). This data is used to proxy the nature of the social interactions in the forms of e.g. dyads and triads (Gaito et al. 2017). Furthermore, Eagle et al. (2009) showed the potential to use CDR to compute several social network metrics such as egodensity, the number of existing edges (links) to the number of possible edges, as well as average tie strength by the volume of calls per degree, contacts. In other words, CDR could be used, similar to human mobility studies, to proxy social networks and social capital by measuring the diversity of calls and the volume of calls per contacts (Castillo et al. 2018; Eagle et al. 2009, 2010; Mamei et al. 2018). Furthermore, some researchers also used the temporal patterns of calls from CDR to measure tie strength (Singh and Ghosh 2017). Other researchers have used the minimum number of calls between nodes to measure the direct influence of nodes in a social network (Peng et al. 2017a, b).

5 What are the key findings for economic geography?

As mentioned previously, studies using mobile phone data cover a variety of scientific disciplines. With regard to the focus of this paper on the geography of connectivity, the insights of mobile phone data concern (1) human mobility on an individual level, (2) individual social networks, and (3) aggregate patterns of human mobility and social networks.

5.1 Human mobility

The research on mobility and activity space relates to seminal work of e.g. Torsten Hägerstrand (1970) on time-geography, which partly set the framework for this school of research (see e.g., Järv et al. 2014; Puura et al. 2018; Yang et al.

2016; Yuan and Raubal 2016). Not only is an individual's activity space moulded by an interdependent relationship between time and space, but it is also shaped by the social and environmental structures of their surroundings. It is influenced by the habits, culture and needs of the individuals and therefore contains not only variance across space but also between individuals (Järv et al. 2014, 2015).

Studies on individual mobility using mobile positioning data identified some regular patterns that hold across cultures and spatial contexts (González et al. 2008; Kang et al. 2012; Yuan and Raubal 2014). In particular, that activity space declines with distance, meaning that most individuals regardless of local context tend to have their activity space within a relatively close distance to their home. This distance decay pattern has been modelled with a power law distribution (González et al. 2008; Moyano et al. 2012; Shi et al. 2017), or an exponential law distribution for intra-urban travel patterns (Kang et al. 2012). In line with this, the radius of gyration—that measures the diversity of places visited—follows the same pattern (González et al. 2008). These findings on the regularity of mobility patterns implies that individuals' activity spaces can be captured within a short timeframe and consequently predicted (Song et al. 2010).

However, despite general regularities, individuals' mobility patterns also vary between cultures and spatial structures (see e.g., Ahas et al. 2015; Järv et al. 2015). Ahas et al. (2015) compared mobility patterns between Paris (France), Tallinn (Estonia) and Harbin (China) and concluded that the different mobility patterns of individuals were consistent with the different economic structures of these cities. Similar findings were provided by Amini et al. (2014) who discovered that the population in Portugal had a much wider activity space, commuting more often and longer than the population in Côte d'Ivoire. Similarly, Yadav et al. (2014) found the total travel distance of urban inhabitants to be around six times lower in a developing country as compared to a developed economy. Such findings point out the importance of an informed understanding of local geographical processes that influence such patterns.

Beyond finding regularities in human mobility patterns and that the local context plays a role in the observed patterns of human mobility there has been research that utilizes mobile positioning data to see how these regular patterns of individual mobility adapt to changes in the local context. An example would be to observe the effects of sudden changes in the environment, due to e.g. disasters or infrastructure changes (Barbosa et al. 2018; Pappalardo and Simini, 2018; Simini et al. 2012). An example is the use of mobile phone data to track changes in travel patterns after plant closure (Toole et al. 2015). Such estimations typically require the creation of models for simulating human mobility from mobile phone data (see e.g., Barbosa et al. 2018; Doyle et al. 2019; Li et al. 2019; Pappalardo and Simini 2018). Yet, the main challenge is to account for the observed variance between individuals and observed deviations from the main mobility patterns, which require both an understanding of regular patterns before the event and data that captures the effects of the event on the regularity of patterns (Pappalardo and Simini 2018).

5.2 Individuals' social networks

Findings on individuals' social networks from mobile phone data relate to three categories: (1) the connection between social networks and human mobility, (2) the differences between spatial contexts, and (3) its connection to socio-economic variables.

A few studies investigate the relationship between an individual's social network and individual mobility. Moyano et al. (2012), Phithakkitnukoon et al. (2012), and Puura et al. (2018) observed a strong relationship between the structure of individuals' social networks and their spatial mobility. Moyano et al. (2012) observed that individuals who call frequently and have a wider range of contacts also tend to have a social network that stretched larger distances. Puura et al. (2018) found that the width of a social network closely followed the range of a person's activity space. If a person has a large social network across a larger geographical distance, its spatial mobility also tended to be higher. This relationship changed between regions as one moved across the regional hierarchy. Larger cities saw a strong relationship between individual's social network and their spatial mobility whilst this correlation became weaker for smaller regions. A cause may be the stronger commuting patterns in large cities as compared to smaller regions (Puura et al. 2018).

As regards differences between spatial contexts, Eagle et al. (2009) find that variations in calling patterns between the capital, urban and rural areas supported a more diversified personal network in urban areas. This difference was explained by behavioural adaptation when moving to urban areas. Similarly, Mamei et al. (2018) used communication data from CDR to accurately proxy social capital of regions in Italy. The findings suggest that regions which overall have a high level of communication within themselves also have stronger social capital, such as association density, whilst communication between areas are negatively correlated with social cohesion of the region. Similar studies were made by Singh and Ghosh (2017) who inferred social capital by using a small dataset of CDR (55 observations), which was joined with phone surveys the users had to fill out. In this study, bridging and bonding social capital could be related to the communication patterns available in CDR.

Lastly, Fixman et al. (2016) found that the social network of users could accurately predict income levels. This was done by using bank information for a subset of the population captured by CDR. Accordingly, caller and callee had a strong tendency towards the same income level and that the amounts of calls followed similar patterns among income levels. The authors inferred that income levels of users could be predicted with 71% accuracy by using the amounts of calls as a predictor. Toole et al. (2015) showed how calling and mobility patterns changed for individuals that experienced employment shocks due to a large lay-off. Not surprisingly, communication and mobility contracted significantly. Rather unique to this study was the use of the measure 'churn' that measured the fraction of contacts that was not called the month afterwards, which significantly increased after the closing event.

5.3 Aggregate patterns of human mobility and social networks

Aggregate patterns of human mobility or social networks are usually analysed together with other regional datasets to investigate regional development or segregation. In particular, the use of mobile phone data in longitudinal studies is considered powerful (Cottineau and Vanhoof 2019; Eagle et al. 2010; Mao et al. 2015; Šćepanović et al. 2015; Schmid et al. 2017). However, there are overall relatively few studies linking such datasets. The main difficulty relates to accessing socio-economic data with a high spatial resolution together with mobile phone data that spans multiple regions (Eagle et al. 2010).

Overall, aggregate patterns of human mobility and social networks exhibit relatively strong associations with regional development. This is mainly used in emerging economies to circumvent the often lacking or outdated socio-economic data available by using mobile phone data as a proxy. Studies conducted by Joshua Blumenstock concluded that the estimations of socio-economic variables using CDR was about as accurate as 5-year old household surveys (Blumenstock 2018; Blumenstock et al. 2015; Blumenstock and Eagle 2010). Yet, the quality of predictions varies by country and models trained in one country do not necessarily provide good fits in another country.

More concretely, Schmid et al. (2017) found that mobile phone data could accurately estimate the literacy rate in Senegal. In South America, Frias-Martinez et al. (2012) found that the socio-economic level of an area was correlated with the radius of gyration, the diversity of visited BTS and the diameter of the area of influence in individual mobility. Wang and Kilmartin (2014) confirmed that mobility patterns in Uganda were a good indicator for regional development. Additionally, they found strong connectivity between the larger and more developed cities. In a similar vein, Šćepanović et al. (2015) found that the more developed regions in Côte d'Ivoire were strong commuting centres but also had a much smaller radius of gyration than poorer regions, reflecting that individuals in poorer regions faced a much larger commute. Furthermore, communication patterns are strong predictors of poverty and education rates (Castillo et al. 2018; Mao et al. 2015).

In the European context, only a few articles have been identified that deal with regional development by using mobile phone data. The most popular is the article in 'Science' by Eagle et al. (2010), which showed that the calling patterns between regions provided an accurate picture of regional development in the UK. Based on this finding, the authors argue that social network diversity increases social and economic opportunities. Similarly, Pappalardo et al. (2015) found that mobility diversity was positively correlated with socio-economic development in French municipalities while no significant relationship was detected for mobility volume. Using calling and mobility patterns within CDR, Bajardi et al. (2015) observed that the spatial cohesion of international communities in Milan was correlated with their income. The less clustered and cohesive the communities were, the better socio-economic status they had.

A set of studies used mobile phone data to assess questions of segregation, which leaves a clear trail in communication and mobility patterns (Hedman et al. 2021; Östh et al. 2018). For instance, Cottineau and Vanhoof (2019) found that mobility

range and diversity tended to decrease in cities with large levels of segregation. Järvi et al. (2015) and Silm et al. (2018) found significant differences in mobility patterns between Estonians and Russians in Tallinn linked to the segregation and lack of integration of the Russian minority. Russians had a much smaller activity space in comparison to Estonians, but this difference reduced with age, indicating that ethnic segregation is larger at younger ages. Finally, Östh et al. (2018) used the trajectory data extracted from mobile phone data to study how spatial mobility shape segregation level, finding that connectivity across space mediated segregation levels and concluded that mere residential location cannot singlehandedly account for socio-economic segregation. Such findings point to the significance of connectivity in regional studies.

6 Challenges and limitations

When handling mobile phone data there are a set of challenges and limitations that are generic across approaches, which make it prone to bias and error (Table 2). We group the challenges and limitations by (1) how the datasets are generated, (2) how the data is processed, and (3) how the analysis is connected to theory.

First, the *data* is not generated primarily for analytical purposes within scientific fields, although certain exceptions exist where the positioning data has been generated with the purpose of research (Ahas et al. 2007a, b; Ahas et al. 2010), but for commercial purposes with the aim to collect information about customers and not about the population. This entails that it reflects a non-random sampling of the population. Especially if data from only one operator is used the question of representability is highly relevant since the dataset will be influenced by factors such as market share, individual preferences for firms and overall competition when it comes to who uses the operator. Lazer and Radford (2017) and Iovan et al. (2013) remind us that this is a source bias due to the nonrandom sampling of the population. However, mobile phone data can, despite its nature, represent sub-groups if it has data from quite well, simply due to the large volume of users and usage within the dataset (Arai et al. 2016; Becker et al. 2011).

Furthermore, ownership bias also plays a role in some CDR datasets. Whilst the technology has diffused to nearly every individual in European countries causing this bias to have reduced in significance, there are still places in the world where socio-economic status and cultural norms create a bias to who is observed in the dataset (Arai et al. 2016). Arai et al. (2016) suggest that a way out is through combining CDR with secondary data that informs us of local contexts and could work as reference points.

Another challenge regarding the nature of mobile phone data relates to spatial and temporal scarcity in the dataset. This is due to the interval between events, calls or text, and within the datasets in which an individual can have passed through several Voronoi polygons without being recorded to do so. The main issue is that an infrequent user could have conducted several trips and activities across an urban environment in the time-period between two events (Zhao et al. 2016). Furthermore, real trajectories do not necessarily follow a direct linear movement between

Table 2 Limitations of mobile phone data

	Limitations	Key issues	Main consequences
Data	Bias	Ownership bias(Arai et al. 2016); Age Income Gender Urban–Rural bias in BTS density (Vanhoof et al. 2018) Individual variance in calling patterns (Puura et al. 2018) Space specific calling patterns (Zhao et al. 2016) Commercially generated data (Lazer and Radford 2017)	Questions the representability for larger communities and regions Variance inflation in predictions Skews the human mobility pattern to emphasise specific places
	Source	Generated only when used (Chen et al. 2018) Infrequent users leaving only small traces of their daily rhythm (Zhao et al., 2016)	Cannot achieve random sampling of the population (Iovan et al. 2013) Usage of the phone not necessarily align with travel behaviour (Zhao et al. 2016)
	Quality	Overlapping BTS radius Nearest tower is not necessarily picked (Rodriguez-Carrion et al. 2018) Varying detail of mobility within the datasets (Vanhoof et al. 2018) Spatial difference in BTS density	Variance inflation in mobility and calling patterns (Cottineau and Vanhoof 2019; Vanhoof et al. 2018) Urban bias in data quality for human mobility
Processing	Time	Large datasets in need of multiple processing steps (De Montjoye et al. 2016)	Time consuming processes that can prove costly for research
	Accuracy	Steps for pseudo-anonymisation (De Montjoye et al. 2013) Data cleaning Data cleaning and calculation moves data further away from the ground truth (Iovan et al. 2013; Lazer and Radford 2017; Zhao et al. 2016) Aggregation of data to appropriate level increasingly deals with MAUP (Cottineau and Vanhoof 2019)	Increasing difficulties in terms of transparency Caution needs to be applied to the strengths of the conclusions that can be made from the analysis
Theoretical connections	Theoretical and methodological understanding	Research mainly done by computer scientists (Lazer and Radford 2017) Lack of comprehensive knowledge of underlying processes in both theory and methodology (Lazer and Radford 2017; Pappalardo et al. 2015)	Tendency to descriptive correlation analysis in socio-economic connections

the locations indicated by the CDR but would spend different times at each area depending on the trajectory (Hoteit et al. 2014). As such, it will generate an incomplete and scarce dataset that is prone to over- or underestimation of mobility indices (Zhao et al. 2016). Chen et al. (2018) found that the completion of individuals' mobility varied between 37% for infrequent users to 80% for frequent users. Lu et al. (2017) found that the radius of gyration in particular was prone to underestimation and Zhao et al. (2016) found that it tended mostly to underestimate the content of an individual's activity space, the diversity of its travel and travel volume. Chen et al. (2018) found this limitation to be especially important when dealing with a dataset that only spanned a short period of time.

Research that has been dealing with these issues around spatial and temporal scarcity found that some weaknesses could be addressed through trajectory reconstructions (Chen et al. 2019; Liu et al. 2018). Often this is done through machine learning processes, and models of movements that take into account the spatial structure and characteristics of individuals (Chen et al. 2018; Hoteit et al. 2014). Sedentary users would for example be better modelled with linear interpolations whilst commuters across larger distances were better described by cubic interpolations. Another solution has been proposed that used the temporal patterns found in the sparse dataset to interpolate the mobile positioning data of users (Chen et al. 2019; Hoteit et al. 2014). However, beyond positioning and accuracy of activity spaces, the individual's presence in the dataset is due to its calling frequency, and thus the nature of its social network, creates a challenge in of itself. Iovan et al. (2013) argued that it could create serious issues since the mobility patterns vary with calling frequencies such that mobility patterns obtained from frequent callers cannot accurately estimate patterns for infrequent callers. The mobile phone locational data that records a phone's location whenever it pings a BTS (e.g. for internet access or data transfer through apps and not only for calls or text messages) decreases this problem. Consequently, the issue of the infrequent caller might be solved as mobile phone technology increasingly connect phones with BTS even in passive use.

Finally, a significant—especially in the field of economic geography—challenge when using mobile phone data arises from how it is generated when connecting to BTS (see e.g., Batran et al. 2018; Chen et al. 2018; Vanhoof et al. 2018). Since the BTSs are rarely placed equally across space, and frequently follow existing urban structures with high concentrations in urban cores and low density in the rural peripheries, there are spatial bias in accuracy of human mobility and consequently limitations in linking calling patterns with local socio-economic conditions. This means that there are vast differences in detail of human mobility as scale increases (Vanhoof et al. 2018).

Furthermore, the coverage of BTS, especially in urban settings, overlaps with each other. This creates an issue for mobility analysis due to the nature of how phones connect to BTS. They do not necessarily connect to the nearest BTS, instead the decision is influenced by other factors such as the existing usage of each BTS. The issue according to Rodriguez-Carrion et al. (2018) arises due to a ping-pong phenomenon where a phone can go between neighbouring BTS without necessarily having moved, creating a false movement in the dataset. This issue becomes more important when a phone frequently contacts these BTS. This would imply that

equivalent travel behaviours would have different data generated. This will introduce increased variance in travel distance among users, increased differences in activity space for e.g. similar users and it will also lead to uncertainties in the mobility entropies, not only between regions but within a region's users (Batra et al. 2018; Vanhoof et al. 2018). Vanhoof et al. (2018) concluded that existing mobility entropy correlated strongly with the density of BTS and would therefore be unsuitable when comparing regions that have vastly different structure and density of BTS. To address this, they suggested a corrected mobility entropy where the density of BTS would work as a weight for the mobility entropy that would correct the vast difference in density between regions.

Second, challenges relate to the *processing* of mobile phone data and linking it to other datasets. The challenge varies between studies as it relates to the size of the datasets and the nature of the datasets it would be joined with. Some studies have only focused on the Voronoi tessellation that has been generated by the BTS coordinates and not aggregated it further (e.g., Lu et al. 2017; Vanhoof et al. 2018), whilst others have had to deal with the issues of aggregating these Voronoi tessellation to fit administrative spatial units for which complementary data was collected (e.g., Cottineau and Vanhoof 2019; Pappalardo et al. 2015). The aggregation to Voronoi tessellation and then administrative units constitutes a Modifiable Areal Unit Problem (MAUP) since neither of these geographical units necessarily correspond to the nature of the empirical phenomena of interest (Cottineau and Vanhoof 2019; Vanhoof et al. 2018). MAUP is a statistical biasing effect that stems from aggregating point data through arbitrarily defined spatial zoning systems (Hall et al. 2004). This causes some concerns when inferring from observed patterns and values to potential causal relationships.

As regards social network analysis, the challenges relate one hand to the sheer amount of data in CDR datasets and to what this dataset actually reflects, and, on the other hand, they relate to density of BTS towers to join with local socio-economic conditions. Regarding the former, Puura et al. (2018) highlight that CDR do not contain any qualitative information on the nature of the contacts. Using data such as duration of calls to measure tie strengths can become problematic since studies have found that the duration of calls is influenced by the distance between individuals. Moyano et al. (2012) identified that the calls tended to last longer if the geographic distance between contacts was greater. Therefore, when studying social networks Karikoski and Nelimarkka (2011) advised that one should use multiple datasets to more accurately infer the social network of individuals. This however increases complexity because it would require more extensive processing of the data (Lazer and Radford 2017; Raento et al. 2009).

Finally, an important challenge in dealing with mobile phone data in both social network analysis and mobility studies is the issue of privacy and research ethics. This concerns the ability to track users in space-time where some argue that it can be relatively easy to identify individuals from what is believed to be anonymous CDR (de Montjoye et al. 2013), which creates challenges when handling the data. The details of mobile phone data puts pressure on creating an anonymous dataset and representation of the observed patterns to adhere to individual's integrity. The issue is strongly related to the spatio-temporal resolution and the number of

observations of a single phone. The more detailed spatial scale and the more frequent temporal monitoring over a longer duration of observation the larger the risk that individuals can be identified. As such, the chosen resolution of the dataset influences the sensitivity of the data and the need of ethical considerations.

The challenges outlined above reveal that an analysis of mobile phone data requires a significant investment and careful handling, but that there are solutions to some of these challenges. The process of handling and analysing mobile phone data is a time-consuming and costly procedure that puts limitations on research projects (de Montjoye et al. 2016), but continued research on mobile phone data holds potential in providing increasing variety in processes to handle this data. Whilst the potential for the various types of biases discussed above, and their possible consequences for observed patterns, need to be examined in order to ensure the validity of interpretations and conclusions (Calabrese et al. 2013; Chen et al. 2018; Zhao et al. 2016), there has been research produced that have found ways to minimize the effects of some (e.g. Chen et al. 2019; Vanhoof et al. 2018).

Thirdly as regards the *theoretical connections* of analyses of mobile phone data, the lack of social scientists in the field is a limitation (Lazer and Radford 2017). A majority of the works that were found through our systemic literature review was produced by or in collaboration with researchers in fields such as computer or information sciences. Whilst there has been some works produced within the fields of social science, such as Ahas et al. (2015), Eagle et al. (2010), Hedman et al. (2021), and Östh et al. (2018), they are a minority in our sample. Nevertheless, these researchers have produced insights in a number of social phenomena with relatively straightforward incorporation of mobile phone data in the methodological framework. Essentially, this points towards both a potential for social science and a limitation in the existing research that has been produced. On the one hand there could be low-hanging fruits in the study of networks or human mobility that would easily build on traditional literature. On the other hand, the cumulative production of research that are done by researchers that lack training in the social sciences can build a critical mass towards a science of society rather than exploring the breadth and variety of knowledge that mobile phone data, and big data in general offers (Lazer et al. 2009; Weinhardt, 2021).

7 Concluding discussion

Mobile phone service providers worldwide have access to data from almost eight billion mobile phone subscribers, a number that almost doubled over the last ten years. An estimated 95% of the inhabited world has at least second-generation (2G) cell phone coverage. That makes mobile phone data an unprecedented rich source for the study of human mobility and interactions with both people and space. Yet, its integration in human geography in general, and economic geography specifically is basically lacking from mainstream research. Our findings point to several attempts to incorporate mobile phone data in different ways within current debates in human geography (e.g. in tourism studies), where it holds a rich potential for time geography, connectivity, and development studies. However,

the large contributions from sciences outside the field of human geography opens a gap for fields such as economic geography to utilise mobile phone data.

Mobile phone data are mainly used in two distinct ways. In smaller studies, sensing data generated automatically with the use of the phone can be combined with, for example, phone surveys and proximity measures facilitated through Bluetooth or Wi-Fi related techniques. This is an extension and complement to time-use methodologies gathering data on “what” people do, where they do it (location), and who else is involved (social networks). This type of method involves consent from the user and involves some downloading and installation of additional software on the phone device to answer questions related to “what” people do. It is rare with surveys of this type to extend beyond some hundreds of respondents.

The other approach largely disregards the question of “what” and focusing on location and the massive volume of observations on mobility patterns. From few variables, human mobility can be traced over time and space. The spatial precision varies as a function of mast density, which can be rather high in urban areas and low in rural areas. The more recent form of mobile phone location data, which pings masts regularly irrespectively of calling or SMS activity, provides a rich basis to create anchor points such as home location, work. Typical derived mobility measures are travel distance (as straight lines between masts), range of activity space and heterogeneity of travels.

Such mobile phone data has been used to answer questions about human mobility, social networks, and how aggregate patterns of human mobility and social networks relate to socio-economic development. Even though the results of these studies are highly relevant for economic geography, a major drawback is the lack of theorizing (Pappalardo et al. 2015). The majority of studies have been conducted by computer scientists who are not necessarily trained in human or economic geography (Lazer and Radford, 2017). As a consequence, research using mobile phone data has the tendency of being descriptive, focussing on correlations, and of being scarcely related to literature and theory in economic geography.

This is a missed opportunity for several reasons. First, the finding that human mobility follows general patterns is important for economic geography, which operates with the assumption that knowledge exchange has a distance decay, meaning that the likelihood to exchange knowledge decreases with distance. However, there is hardly any empirical evidence on how the distance decay looks like. Articles in economic geography use scarce sources on business travels to infer on the distance decay (Andersson and Karlsson 2007; Grillitsch and Nilsson 2015). Mobile phone data provide an extremely rich source to empirically unveil the distance decay based on human mobility and calibrate it for different regional contexts (i.e. mobility patterns differ between larger cities and rural areas).

Second, the finding that human mobility and communication patterns are closely related to social networks and social capital ties into a hot topic in economic geography (e.g. Cortinovis et al. 2017; Ettlinger 2003; Giuliani, 2007; Kemeny et al. 2016). However, data on social networks are limited to patent or publication data, or rather small-scale surveys. Furthermore, the intangible aspects of social capital related to institutions and trust are hard to measure (Rodríguez-Pose 2013). Measures derived

from mobile phone data could be mobilized as a potentially strong proxy for social networks and social capital in regions.

Third, the finding that human mobility and communication patterns exhibit a strong correlation with economic development in regions supports the very basic assumptions that connectivity is a fundamental factor in explaining the geography of the economy. Yet, the measures of connectivity in economic geography are rather limited and encompass the mentioned patent and publication data, as well as information from surveys such as the Community Innovation Survey in the European Union. These data are scarce in time (e.g. yearly register data or event-based data on publications or patents) and often scarce in scale (e.g. uses data on regions, metropolitan areas, or nations).

In contrast, mobile phone data have a high resolution in time and in space and can therefore yield powerful and complementary insights about the connectivity within and between regions, as well as how connectivity links to different patterns of economic development. This would require the combination of mobile phone data with other socio-economic data, thereby adding measures of connectivity to structural factors influencing regional development. Even though such an analysis may not be fine grained enough for causal inference, it would allow investigating patterns of correlations and assessing whether they are in line with theoretically derived hypotheses linking regional development to connectivity. The potential for causal inferences increases if mobile phone data can be matched closer “to the individual”, i.e. at a high spatial resolution. This is possible by combining mobile phone data with grid-level data (e.g. geo-coded data on small squares), which is becoming increasingly available. Furthermore, mobile phone data has been available since the 2000s potentially allowing for longitudinal studies, which also increase the potential for causal inference. Even if just available for a specific point in time, mobile phone data hold value as a cross-sectional data representative for specific scales which have been used by e.g. Östh et al. (2018).

Adding data on people’s location in five-minutes interval and their communication network can enhance our understanding about connectivity and its role in economic development in space but there are methodological issues to resolve. The problem at hand is one of combining rich data sources with scarce data sources. The case could typically be several hundreds of thousands of mobile phone data observations (more in time than space) to be associated with household statistics or register data (many attributes but once a year). In cases when data is available at individual or household level it is a question of aggregating this data to Voronoi polygons. When socio-economic data is available on aggregated levels such as municipalities or regions then the opposite process is necessary—Voronoi polygons need to be aggregated to a region. It is also possible to downscale data based on some known relation, e.g. social network characteristics and income levels. This has not been addressed by the research literature to the best of our knowledge.

Another issue relates to the three main sources of bias that mobile phone data are susceptible to. Firstly, aggregation errors may occur during multiple stages of the process from collection to analysis. This relates to the abstraction of mast coverage using Voronoi-polygons and when that data is combined with other datasets. Aggregation can also raise issues related to ecological fallacy, especially moving between

scales where especially the findings from groups are at risk of being extrapolated to say something about individuals. Secondly, a sampling bias may occur due to the large variation between individuals in calling volume. This has been addressed by inferring patterns from other observations or by using a wider time frame together with an exclusion of infrequent users. Lastly, the varying density of mast towers follows the spatial structure of regions. This provides variance in the observed patterns, which inflates the movements of urban users and reduces the movements of rural users. A proposed solution is to weigh measures by the density of mast towers.

Authors' contribution Data collection, A.E.; Analysis, A.E.; Project administration, M.G. and O.H.; Supervision, M.G. and O.H.; Validation, M.G. and O.H.; Writing—original draft, A.E.; Writing—review and editing, AE, M.G. and O.H. All authors have read and agreed to the final version of the manuscript.

Funding Open access funding provided by Lund University. Not applicable.

Availability of data and material A registry on the reviewed articles is available through correspondence with the authors.

Code availability Not applicable.

Declarations

Conflict of interest Not applicable.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Ahas R, Aasa A, Mark Ü, Pae T, Kull A (2007a) Seasonal tourism spaces in Estonia: case study with mobile positioning data. *Tour Manag* 28(3):898–910. <https://doi.org/10.1016/j.tourman.2006.05.010>
- Ahas R, Aasa A, Silm S, Aunap R, Kalle H, Mark Ü (2007b) Mobile positioning in space-time behaviour studies: social positioning method experiments in Estonia. *Cartogr Geogr Inf Sci* 34(4):259–273. <https://doi.org/10.1559/152304007782382918>
- Ahas R, Aasa A, Silm S, Tiru M (2010) Daily rhythms of suburban commuters' movements in the Tallinn metropolitan area: case study with mobile positioning data. *Transp Res Part C Emerg Technol* 18(1):45–54. <https://doi.org/10.1016/j.trc.2009.04.011>
- Ahas R, Aasa A, Yuan Y, Raubal M, Smoreda Z, Liu Y, Ziemiłcki C, Tiru M, Zook M (2015) Everyday space–time geographies: using mobile phone-based sensor data to monitor urban activity in Harbin, Paris, and Tallinn. *Int J Geogr Inf Sci* 29(11):2017–2039. <https://doi.org/10.1080/13658816.2015.1063151>

- Amini A, Kung K, Kang C, Sobolevsky S, Ratti C (2014) The impact of social segregation on human mobility in developing and industrialized regions. *EPJ Data Sci* 3(1):1–20. <https://doi.org/10.1140/epjds31>
- Andersson M, Karlsson C (2007) Knowledge in regional economic growth—the role of knowledge accessibility. *Ind Innov* 14(2):129–149. <https://doi.org/10.1080/13662710701252450>
- Arai A, Fan Z, Matekenya D, Shibasaki R (2016) Comparative perspective of human behavior patterns to uncover ownership bias among mobile phone users. *ISPRS Int J Geo-Inf* 5(6):85. <https://doi.org/10.3390/ijgi5060085>
- Bajardi P, Delfino M, Panisson A, Petri G, Tizzoni M (2015) Unveiling patterns of international communities in a global city using mobile phone data. *EPJ Data Sci* 4(1):1–17. <https://doi.org/10.1140/epjds/s13688-015-0041-5>
- Barbosa A, Barthelemy M, Ghoshal G, James CR, Lenormand M, Louail T, Menezes R, Ramasco JJ, Simini F, Tomasini M (2018) Human mobility: models and applications. *Phys Rep* 734:1–74. <https://doi.org/10.1016/j.physrep.2018.01.001>
- Bathelt H, Malmberg A, Maskell P (2004) Clusters and knowledge: local buzz, global pipelines and the process of knowledge creation. *Prog Hum Geogr* 28(1):31–56. <https://doi.org/10.1191/0309132504ph469oa>
- Batran M, Mejia MG, Kanasugi H, Sekimoto Y, Shibasaki R (2018) Inferencing human spatiotemporal mobility in greater maputo via mobile phone big data mining. *ISPRS Int J Geo-Inf* 7(7):259. <https://doi.org/10.3390/ijgi7070259>
- Becker RA, Caceres R, Hanson K, Loh JM, Urbanek S, Varshavsky A, Volinsky C (2011) A tale of one city: using cellular network data for urban planning. *IEEE Pervasive Comput* 10(4):18–26. <https://doi.org/10.1109/MPRV.2011.44>
- Bettencourt LMA, Lobo J, Strumsky D, West GB (2010) Urban scaling and its deviations: revealing the structure of wealth, innovation and crime across cities. *PLoS ONE* 5(11):e13541. <https://doi.org/10.1371/journal.pone.0013541>
- Blumenstock J (2018) Estimating economic characteristics with phone data. *AEA Pap Proc* 108:72–76. <https://doi.org/10.1257/pandp.20181033>
- Blumenstock J, Eagle N (2010) Mobile divides: gender, socioeconomic status, and mobile phone use in Rwanda. In: Proceedings of the 4th ACM/IEEE international conference on information and communication technologies and development, pp 1–10. <https://doi.org/10.1145/2369220.2369225>
- Blumenstock J, Cadamuro G, On R (2015) Predicting poverty and wealth from mobile phone metadata. *Science* 350(6264):1073–1076. <https://doi.org/10.1126/science.aac4420>
- Calabrese F, Smoreda Z, Blondel VD, Ratti C (2011) Interplay between telecommunications and face-to-face interactions: a study using mobile phone data. *PLoS ONE* 6(7):e20814. <https://doi.org/10.1371/journal.pone.0020814>
- Calabrese F, Diao M, Di Lorenzo G, Ferreira J, Ratti C (2013) Understanding individual mobility patterns from urban sensing data: a mobile phone trace example. *Transp Res Part C Emerg Technol* 26:301–313. <https://doi.org/10.1016/j.trc.2012.09.009>
- Castillo G, Layedra F, Guaranda M-B, Lara P, Vaca C (2018) The silence of the cantons: estimating villages socioeconomic status through mobile phones data. In: 2018 International conference on e-democracy e-government (ICEDEG), pp 172–178. <https://doi.org/10.1109/ICEDEG.2018.8372308>
- Chen G, Hoteit S, Viana AC, Fiore M, Sarraute C (2018) Enriching sparse mobility information in call detail records. *Comput Commun* 122:44–58. <https://doi.org/10.1016/j.comcom.2018.03.012>
- Chen G, Viana AC, Fiore M, Sarraute C (2019) Complete trajectory reconstruction from sparse mobile phone data. *EPJ Data Sci* 8(1):30. <https://doi.org/10.1140/epjds/s13688-019-0206-8>
- Chi G, Thill J-C, Tong D, Shi L, Liu Y (2016) Uncovering regional characteristics from mobile phone data: a network science approach. *Pap Reg Sci* 95(3):613–631. <https://doi.org/10.1111/pirs.12149>
- Cortinovis N, Xiao J, Boschma R, van Oort FG (2017) Quality of government and social capital as drivers of regional diversification in Europe. *J Econ Geogr* 17(6):1179–1208. <https://doi.org/10.1093/jeg/lbx001>
- Cottineau C, Vanhoof M (2019) Mobile phone indicators and their relation to the socioeconomic organisation of cities. *ISPRS Int J Geo-Inf* 8(1):19. <https://doi.org/10.3390/ijgi8010019>
- Dash M, Koo KK, Hollecsek T, Yap G-E, Krishnaswamy SP, Shi-Nash A (2015) From mobile phone data to transport network—gaining insight about human mobility. In: 2015 16th IEEE international conference on mobile data management, vol 1, pp 243–250. <https://doi.org/10.1109/MDM.2015.74>

- Dashdorj Z, Sobolevsky S, Serafini L, Ratti C (2014) Human activity recognition from spatial data sources. In: Proceedings of the third ACM SIGSPATIAL international workshop on mobile geographic information systems, pp 18–25. <https://doi.org/10.1145/2675316.2675321>
- Dashdorj Z, Sobolevsky S, Lee S, Ratti C (2018) Deriving human activity from geo-located data by ontological and statistical reasoning. *Knowl Based Syst* 143:225–235. <https://doi.org/10.1016/j.knosys.2017.11.038>
- de la Roca J, Puga D (2017) Learning by working in big cities. *Rev Econ Stud* 84(1):106–142. <https://doi.org/10.1093/restud/rdw031>
- de Montjoye Y-A, Hidalgo CA, Verleysen M, Blondel VD (2013) Unique in the crowd: the privacy bounds of human mobility. *Sci Rep* 3(1):1376. <https://doi.org/10.1038/srep01376>
- de Montjoye Y-A, Rocher L, Pentland AS (2016) bandicoot: a Python toolbox for mobile phone meta-data. *J Mach Learn Res* 17(175):1–5
- Deville P, Linard C, Martin S, Gilbert M, Stevens FR, Gaughan AE, Blondel VD, Tatem AJ (2014) Dynamic population mapping using mobile phone data. *Proc Natl Acad Sci* 111(45):15888–15893. <https://doi.org/10.1073/pnas.1408439111>
- Doyle C, Herga Z, Dipple S, Szymanski BK, Korniss G, Mladenčić D (2019) Predicting complex user behavior from CDR based social networks. *Inf Sci* 500:217–228. <https://doi.org/10.1016/j.ins.2019.05.082>
- Eagle N, Pentland A, Lazer D (2008) Mobile phone data for inferring social network structure. In: Liu H, Salerno JJ, Young MJ (eds) *Social computing, behavioral modeling, and prediction*. Springer, US, Boston, pp 79–88. https://doi.org/10.1007/978-0-387-77672-9_10
- Eagle N, de Montjoye Y-A, Bettencourt LMA (2009) Community computing: comparisons between rural and urban societies using mobile phone data. In: 2009 International conference on computational science and engineering, vol 4, pp 144–150. <https://doi.org/10.1109/CSE.2009.91>
- Eagle N, Macy M, Claxton R (2010) Network diversity and economic development. *Science* 328(5981):1029–1031. <https://doi.org/10.1126/science.1186605>
- Engelmann G, Smith G, Goulding J (2018) The unbanked and poverty: predicting area-level socio-economic vulnerability from M-money transactions. In: 2018 IEEE international conference on big data (big data), pp 1357–1366. <https://doi.org/10.1109/BigData.2018.8622268>
- Ettlinger N (2003) Cultural economic geography and a relational and microspace approach to trusts, rationalities, networks, and change in collaborative workplaces. *J Econ Geogr* 3(2):145–171. <https://doi.org/10.1093/jeg/3.2.145>
- Farber S, Neutens T, Miller HJ, Li X (2013) The social interaction potential of metropolitan regions: a time-geographic measurement approach using joint accessibility. *Ann Assoc Am Geogr* 103(3):483–504. <https://doi.org/10.1080/00045608.2012.689238>
- Farole T, Rodríguez-Pose A, Storper M (2011) Human geography and the institutions that underlie economic growth. *Prog Hum Geogr* 35(1):58–80. <https://doi.org/10.1177/0309132510372005>
- Fixman M, Berenstein A, Brea J, Minnoni M, Travizano M, Sarraute C (2016) A Bayesian approach to income inference in a communication network. In: 2016 IEEE/ACM international conference on advances in social networks analysis and mining (ASONAM), pp 579–582. <https://doi.org/10.1109/ASONAM.2016.7752294>
- Frias-Martinez V, Virseda-Jerez J, Frias-Martinez E (2012) On the relation between socio-economic status and physical mobility. *Inf Technol Dev* 18(2):91–106. <https://doi.org/10.1080/02681102.2011.630312>
- Gaito S, Quadri C, Rossi GP, Zignani M (2017) Urban communications and social interactions through the lens of mobile phone data. *Online Soc Netw Media* 1:70–81. <https://doi.org/10.1016/j.osnem.2017.04.003>
- Gao S, Liu Y, Wang Y, Ma X (2013) Discovering spatial interaction communities from mobile phone data. *Trans GIS* 17(3):463–481. <https://doi.org/10.1111/tgis.12042>
- Ghosh I, Singh VK (2018) Modeling social support scores using phone use patterns. *Proc Assoc Inf Sci Technol* 55(1):133–142. <https://doi.org/10.1002/pr2.2018.14505501015>
- Giuliani E (2007) The selective nature of knowledge networks in clusters: evidence from the wine industry. *J Econ Geogr* 7(2):139–168. <https://doi.org/10.1093/jeg/lbl014>
- Glaeser EL, Maré D (2015) Cities and skills. *J Labor Econ*. <https://doi.org/10.1086/319563>
- González MC, Hidalgo CA, Barabási A-L (2008) Understanding individual human mobility patterns. *Nature* 453(7196):779–782. <https://doi.org/10.1038/nature06958>

- Grillitsch M, Nilsson M (2015) Innovation in peripheral regions: Do collaborations compensate for a lack of local knowledge spillovers? *Ann Reg Sci* 54(1):299–321. <https://doi.org/10.1007/s00168-014-0655-8>
- Hägerstrand T (1970) What about people in regional science? *Pap Reg Sci Associ* 24(1):6–21. <https://doi.org/10.1007/BF01936872>
- Hall O, Hay GJ, Bouchard A, Marceau DJ (2004) Detecting dominant landscape objects through multiple scales: an integration of object-specific methods and watershed segmentation. *Landsc Ecol* 19(1):59–76. <https://doi.org/10.1023/B:LAND.0000018371.43447.1f>
- Hedman L, Kadarik K, Andersson R, Östh J (2021) Daily mobility patterns: Reducing or reproducing inequalities and segregation? *Soc Incl* 9(2):208–221
- Hernandez M, Hong L, Frias-Martinez V, Whitby A, Frias-Martinez E (2017) Estimating poverty using cell phone data: evidence from Guatemala (SSRN scholarly paper ID 2923520). Social Science Research Network. <https://papers.ssrn.com/abstract=2923520>
- Hotëit S, Secci S, Sobolevsky S, Ratti C, Pujolle G (2014) Estimating human trajectories and hotspots through mobile phone data. *Comput Netw* 64:296–307. <https://doi.org/10.1016/j.comnet.2014.02.011>
- Iovan C, Olteanu-Raimond A-M, Couronné T, Smoreda Z (2013) Moving and calling: mobile phone data quality measurements and spatiotemporal uncertainty in human mobility studies. In: Vandenberghe D, Bucher B, Cromptoets J (eds) *Geographic information science at the heart of Europe*. Springer, Berlin, pp 247–265. https://doi.org/10.1007/978-3-319-00615-4_14
- Jacobs J (1969) *The economy of cities*. Random House, New York
- Jahani E, Sundsøy P, Bjelland J, Bengtsson L, Pentland A, de Montjoye Y-A (2017) Improving official statistics in emerging markets using machine learning and mobile phone data. *EPJ Data Sci* 6(1):1–21. <https://doi.org/10.1140/epjds/s13688-017-0099-3>
- Järv O, Ahas R, Witlox F (2014) Understanding monthly variability in human activity spaces: a twelve-month study using mobile phone call detail records. *Transp Res Part C Emerg Technol* 38:122–135. <https://doi.org/10.1016/j.trc.2013.11.003>
- Järv O, Müürisepp K, Ahas R, Derudder B, Witlox F (2015) Ethnic differences in activity spaces as a characteristic of segregation: a study based on mobile phone usage in Tallinn, Estonia. *Urban Stud* 52(14):2680–2698. <https://doi.org/10.1177/0042098014550459>
- Jiang S, Ferreira J, González MC (2012) Clustering daily patterns of human activities in the city. *Data Min Knowl Discov* 25(3):478–510. <https://doi.org/10.1007/s10618-012-0264-z>
- Kang C, Ma X, Tong D, Liu Y (2012) Intra-urban human mobility patterns: an urban morphology perspective. *Physica A Stat Mech Appl* 391(4):1702–1717. <https://doi.org/10.1016/j.physa.2011.11.005>
- Karikoski J, Nelimarkka M (2011) Measuring social relations with multiple datasets. *Int J Soc Comput Cyber Phys Syst* 1(1):98–113. <https://doi.org/10.1504/IJSCCPS.2011.043606>
- Kemeny T, Feldman M, Ethridge F, Zoller T (2016) The economic value of local social networks. *J Econ Geogr* 16(5):1101–1122. <https://doi.org/10.1093/jeg/lbv043>
- Kung KS, Greco K, Sobolevsky S, Ratti C (2014) Exploring universal patterns in human home-work commuting from mobile phone data. *PLoS ONE* 9(6):e96180. <https://doi.org/10.1371/journal.pone.0096180>
- Kwan M-P (1999) Gender and individual access to urban opportunities: a study using space-time measures. *Prof Geogr* 51(2):210–227. <https://doi.org/10.1111/0033-0124.00158>
- Kwan M-P (2007) Mobile communications, social networks, and urban travel: hypertext as a new metaphor for conceptualizing spatial interaction. *Prof Geogr* 59(4):434–446. <https://doi.org/10.1111/j.1467-9272.2007.00633.x>
- Lazer D, Radford J (2017) Data ex machina: introduction to big data. *Annu Rev Sociol* 43(1):19–39. <https://doi.org/10.1146/annurev-soc-060116-053457>
- Lazer D, Pentland A, Adamic L, Aral S, Barabási A-L, Brewer D, Christakis N, Contractor N, Fowler J, Gutmann M, Jebara T, King G, Macy M, Roy D, Van Alstyne M (2009) Computational social science. *Science* 323(5915):721–723. <https://doi.org/10.1126/science.1167742>
- Li M, Gao S, Lu F, Zhang H (2019) Reconstruction of human movement trajectories from large-scale low-frequency mobile phone data. *Comput Environ Urban Syst* 77:101346. <https://doi.org/10.1016/j.compenvurbysys.2019.101346>
- Liu Z, Ma T, Du Y, Pei T, Yi J, Peng H (2018) Mapping hourly dynamics of urban population using trajectories reconstructed from mobile phone records. *Trans GIS* 22(2):494–513. <https://doi.org/10.1111/tgis.12323>

- Lu S, Fang Z, Zhang X, Shaw S-L, Yin L, Zhao Z, Yang X (2017) Understanding the representativeness of mobile phone location data in characterizing human mobility indicators. *ISPRS Int J Geo-Inf* 6(1):7. <https://doi.org/10.3390/ijgi6010007>
- Mamei M, Pancotto F, De Nadai M, Lepri B, Vescovi M, Zambonelli F, Pentland A (2018) Is social capital associated with synchronization in human communication? An analysis of Italian call records and measures of civic engagement. *EPJ Data Sci* 7(1):1–16. <https://doi.org/10.1140/epjds/s13688-018-0152-x>
- Manfredini F, Pucci P, Tagliolato P (2013) Deriving mobility practices and patterns from mobile phone data. In: Murgante B, Misra S, Carlini M, Torre CM, Nguyen H-Q, Taniar D, Apduhan BO, Gervasi O (eds) *Computational science and its applications—ICCSA 2013*. Springer, Berlin, pp 438–451. https://doi.org/10.1007/978-3-642-39646-5_32
- Mao H, Shuai X, Ahn Y-Y, Bollen J (2015) Quantifying socio-economic indicators in developing countries from mobile phone communication data: applications to Côte d'Ivoire. *EPJ Data Sci* 4(1):1–16. <https://doi.org/10.1140/epjds/s13688-015-0053-1>
- Martin R, Aslesen HW, Grillitsch M, Herstad SJ (2018) Regional Innovation systems and global flows of knowledge. In: Isaksen A, Martin R, Trippel M (eds) *New avenues for regional innovation systems—theoretical advances, empirical cases and policy lessons*. Springer, Berlin, pp 127–147. https://doi.org/10.1007/978-3-319-71661-9_7
- Mota T, Munjal A, Camp T (2015) Large-scale human mobility analysis based on mobile phone and social media communication: a case-study in Africa. In: 2015 16th IEEE international conference on mobile data management, vol 2, pp 86–91. <https://doi.org/10.1109/MDM.2015.84>
- Moyano LG, Thomae ORM, Frias-Martinez E (2012) Uncovering the spatio-temporal structure of social networks using cell phone records. In: 2012 IEEE 12th international conference on data mining workshops, pp 242–249. <https://doi.org/10.1109/ICDMW.2012.132>
- Onnela J-P, Arbesman S, González MC, Barabási A-L, Christakis NA (2011) Geographic constraints on social network groups. *PLoS ONE* 6(4):e16939. <https://doi.org/10.1371/journal.pone.0016939>
- Östh J, Shuttleworth I, Niedomysl T (2018) Spatial and temporal patterns of economic segregation in Sweden's metropolitan areas: a mobility approach. *Environ Plan A Econ Space* 50(4):809–825. <https://doi.org/10.1177/0308518X18763167>
- Pappalardo L, Simini F (2018) Data-driven generation of spatio-temporal routines in human mobility. *Data Min Knowl Discov* 32(3):787–829. <https://doi.org/10.1007/s10618-017-0548-4>
- Pappalardo L, Pedreschi D, Smoreda Z, Giannotti F (2015) Using big data to study the link between human mobility and socio-economic development. In: 2015 IEEE international conference on big data (big data), pp 871–878. <https://doi.org/10.1109/BigData.2015.7363835>
- Peng S, Wang G, Xie D (2017a) Social influence analysis in social networking big data: opportunities and challenges. *IEEE Netw* 31(1):11–17. <https://doi.org/10.1109/MNET.2016.1500104NM>
- Peng S, Yang A, Cao L, Yu S, Xie D (2017b) Social influence modeling using information theory in mobile social networks. *Inf Sci* 379:146–159. <https://doi.org/10.1016/j.ins.2016.08.023>
- Phithakkitnukoon S, Smoreda Z (2016) Influence of social relations on human mobility and sociality: a study of social ties in a cellular network. *Soc Netw Anal Min* 6(1):42. <https://doi.org/10.1007/s13278-016-0351-z>
- Phithakkitnukoon S, Smoreda Z, Olivier P (2012) Socio-geography of human mobility: a study using longitudinal mobile phone data. *PLoS ONE* 7(6):e39253. <https://doi.org/10.1371/journal.pone.0039253>
- Puura A, Silm S, Ahas R (2018) The relationship between social networks and spatial mobility: a mobile-phone-based study in Estonia. *J Urban Technol* 25(2):7–25. <https://doi.org/10.1080/10630732.2017.1406253>
- Raento M, Oulasvirta A, Eagle N (2009) Smartphones: an emerging tool for social scientists. *Sociol Methods Res* 37(3):426–454. <https://doi.org/10.1177/0049124108330005>
- Ratti C, Frenchman D, Pulselli RM, Williams S (2006) Mobile landscapes: using location data from cell phones for urban analysis. *Environ Plan B Plan Des* 33(5):727–748. <https://doi.org/10.1068/b32047>
- Raun J, Ahas R, Tiru M (2016) Measuring tourism destinations using mobile tracking data. *Tour Manag* 57:202–212. <https://doi.org/10.1016/j.tourman.2016.06.006>
- Rodriguez-Carrion A, Garcia-Rubio C, Campo C (2018) Detecting and reducing biases in cellular-based mobility data sets. *Entropy* 20(10):736. <https://doi.org/10.3390/e20100736>
- Rodríguez-Pose A (2013) Do institutions matter for regional development? *Reg Stud* 47(7):1034–1047. <https://doi.org/10.1080/00343404.2012.748978>

- Šćepanović S, Mishkovski I, Hui P, Nurminen JK, Ylä-Jääski A (2015) Mobile phone call data as a regional socio-economic proxy indicator. *PLoS ONE* 10(4):e0124160. <https://doi.org/10.1371/journal.pone.0124160>
- Scherrer L, Tomko M, Ranacher P, Weibel R (2018) Travelers or locals? Identifying meaningful sub-populations from human movement data in the absence of ground truth. *EPJ Data Sci* 7(1):1–21. <https://doi.org/10.1140/epjds/s13688-018-0147-7>
- Schmid T, Bruckschen F, Salvati N, Zbiranski T (2017) Constructing sociodemographic indicators for national statistical institutes by using mobile phone data: estimating literacy rates in Senegal. *J R Stat Soc Ser A (stat Soc)* 180(4):1163–1190. <https://doi.org/10.1111/rssa.12305>
- Scholten C, Friberg T, Sandén A (2012) Re-reading time-geography from a gender perspective: examples from gendered mobility. *Tijdschr Voor Econ Soc Geogr* 103(5):584–600. <https://doi.org/10.1111/j.1467-9663.2012.00717.x>
- Sherman JE, Spencer J, Preisser JS, Gesler WM, Arcury TA (2005) A suite of methods for representing activity space in a healthcare accessibility study. *Int J Health Geogr* 4(1):24. <https://doi.org/10.1186/1476-072X-4-24>
- Shi L, Wang W, Cai W, Wang Z, Zhang S, Zhou W (2017) Mobility patterns analysis of Beijing residents based on call detail records. In: 2017 9th International conference on wireless communications and signal processing (WCSP), pp 1–6. <https://doi.org/10.1109/WCSP.2017.8170955>
- Shoval N, Ahas R (2016) The use of tracking technologies in tourism research: the first decade. *Tour Geogr* 18(5):587–606. <https://doi.org/10.1080/14616688.2016.1214977>
- Silm S, Ahas R, Mooses V (2018) Are younger age groups less segregated? Measuring ethnic segregation in activity spaces using mobile phone data. *J Ethn Migr Stud* 44(11):1797–1817. <https://doi.org/10.1080/1369183X.2017.1400425>
- Simini F, González MC, Maritan A, Barabási A-L (2012) A universal model for mobility and migration patterns. *Nature* 484(7392):96–100. <https://doi.org/10.1038/nature10856>
- Singh VK, Ghosh I (2017) Inferring individual social capital automatically via phone logs. In: Proceedings of the ACM on human–computer interaction, 1(CSCW), vol 95, pp 1–95:12. <https://doi.org/10.1145/3134730>
- Song C, Qu Z, Blumm N, Barabási A-L (2010) Limits of predictability in human mobility. *Science* 327(5968):1018–1021. <https://doi.org/10.1126/science.1177170>
- Steenbruggen J, Tranos E, Nijkamp P (2015) Data from mobile phone operators: A tool for smarter cities? *Telecommun Policy* 39(3):335–346. <https://doi.org/10.1016/j.telpol.2014.04.001>
- Storper M, Venables AJ (2004) Buzz: face-to-face contact and the urban economy. *J Econ Geogr* 4(4):351–370. <https://doi.org/10.1093/jnecg/lbh027>
- Teng W-G, Chou M-C (2007) Mining communities of acquainted mobile users on call detail records. In: Proceedings of the 2007 ACM symposium on applied computing, pp 957–958. <https://doi.org/10.1145/1244002.1244212>
- Thuillier E, Moalic L, Lamrous S, Caminada A (2018) Clustering weekly patterns of human mobility through mobile phone data. *IEEE Trans Mob Comput* 17(4):817–830. <https://doi.org/10.1109/TMC.2017.2742953>
- Toole JL, Lin Y-R, Muehlegger E, Shoag D, González MC, Lazer D (2015) Tracking employment shocks using mobile phone data. *J R Soc Interface* 12(107):20150185. <https://doi.org/10.1098/rsif.2015.0185>
- Tranfield D, Denyer D, Smart P (2003) Towards a methodology for developing evidence-informed management knowledge by means of systematic review. *Br J Manag* 14(3):207–222. <https://doi.org/10.1111/1467-8551.00375>
- Vanhoof M, Schoors W, Van Rompaey A, Ploetz T, Smoreda Z (2018) Comparing regional patterns of individual movement using corrected mobility entropy. *J Urban Technol* 25(2):27–61. <https://doi.org/10.1080/10630732.2018.1450593>
- Wang H, Kilmartin L (2014) Comparing rural and urban social and economic behavior in Uganda: insights from mobile voice service usage. *J Urban Technol* 21(2):61–89. <https://doi.org/10.1080/10630732.2014.888296>
- Weinhardt M (2021) Big data: some ethical concerns for the social sciences. *Soc Sci* 10(2):36. <https://doi.org/10.3390/socsci10020036>
- Werayawarangura N, Pungchaichan T, Vateekul P (2016) Social network analysis of calling data records for identifying influencers and communities. In: 2016 13th international joint conference on computer science and software engineering (JCSSE), pp 1–6. <https://doi.org/10.1109/JCSSE.2016.7748864>

- Wu L, Yang B, Jing P (2016) Travel mode detection based on GPS raw data collected by smartphones: a systematic review of the existing methodologies. *Information* 7(4):67. <https://doi.org/10.3390/info7040067>
- Xu Y, Belyi A, Bojic I, Ratti C (2018) Human mobility and socioeconomic status: analysis of Singapore and Boston. *Comput Environ Urban Syst* 72:51–67. <https://doi.org/10.1016/j.compenvurbsys.2018.04.001>
- Xu Y, Santi P, Ratti C (2021) Beyond distance decay: discover homophily in spatially embedded social networks. *Ann Am Assoc Geogr*. <https://doi.org/10.1080/24694452.2021.1935208>
- Yadav K, Kumar A, Bharati A, Naik V (2014) Characterizing mobility patterns of people in developing countries using their mobile phone data. In: 2014 Sixth international conference on communication systems and networks (COMSNETS), pp 1–8. <https://doi.org/10.1109/COMSNETS.2014.6734892>
- Yang X, Fang Z, Xu Y, Shaw S-L, Zhao Z, Yin L, Zhang T, Lin Y (2016) Understanding spatiotemporal patterns of human convergence and divergence using mobile phone location data. *ISPRS Int J Geo-Inf* 5(10):177. <https://doi.org/10.3390/ijgi5100177>
- Yuan Y, Raubal M (2014) Measuring similarity of mobile phone user trajectories—a spatio-temporal edit distance method. *Int J Geogr Inf Sci* 28(3):496–520. <https://doi.org/10.1080/13658816.2013.854369>
- Yuan Y, Raubal M (2016) Analyzing the distribution of human activity space from mobile phone usage: an individual and urban-oriented study. *Int J Geogr Inf Sci* 30(8):1594–1621. <https://doi.org/10.1080/13658816.2016.1143555>
- Zhao Z, Shaw S-L, Xu Y, Lu F, Chen J, Yin L (2016) Understanding the bias of call detail records in human mobility research. *Int J Geogr Inf Sci* 30(9):1738–1762. <https://doi.org/10.1080/13658816.2015.1137298>
- Zhou C, Huang B-X (2016) Behavioral pattern in mobile phone networks described by the generalized extreme value distribution at individual level. *J Internet Technol* 17(5):1029–1041

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