The Geography of Connectivity:

Trails of Mobile Phone Data

Andreas Erlström (andreas.erlstrom@keg.lu.se)

Department of Human Geography, Lund University, Sweden

Markus Grillitsch (markus.grillitsch@keg.lu.se)

Department of Human Geography & CIRCLE, Lund University, Sweden

Ola Hall (ola.hall@keg.lu.se)

Department of Human Geography, Lund University, Sweden

Papers in Innovation Studies

Paper no. 2020/06

This is a pre-print version of paper,which has been submitted to a journal for review.

Centre for Innovation, Research and Competence in the Learning Economy (CIRCLE)

Lund University

P.O. Box 117, Sölvegatan 16, S-221 00 Lund, SWEDEN

http://www.circle.lu.se/publications

WP 2020/06

The Geography of Connectivity: Trails of Mobile Phone Data

Andreas Erlström, Markus Grillitsch, Ola Hall

Abstract: Connectivity between and within places is one of the cornerstones of human 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 offers a rich and unprecedented source of data, which is exhaustive in time and space closely following movements and partly communication activities of individuals. This paper discusses the state-of-the-art in the analysis of mobile phone data, identifies methodological challenges, elaborates on key findings for geography, and outlines opportunities for future research on the geography of connectivity.

Keywords: Connectivity; mobile phone data; human mobility; social networks; regional development

JEL: B40; J60; O18; R12

The Geography of Connectivity: Trails of Mobile Phone Data

Andreas Erlström[1], Markus Grillitsch[1,2,*], Ola Hall[1]

1) Department of Human Geography, Lund University,

2) CIRCLE – Center for Innovation, Research & Competence in a Learning Economy, Lund University

*corresponding author: markus.grillitsch@keg.lu.se

Abstract:

Connectivity between and within places is one of the cornerstones of human 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 offers a rich and unprecedented source of data, which is exhaustive in time and space closely following movements and partly communication activities of individuals. This paper discusses the state-of-the-art in the analysis of mobile phone data, identifies methodological challenges, elaborates on key findings for geography, and outlines opportunities for future research on the geography of connectivity.

JEL Codes: B40, J60, O18, R12,

Keywords: Connectivity, mobile phone data, human mobility, social networks, regional development

I Introduction

Connectivity between and within places is one of the fundamental cornerstones of geography. Rather than co-location, it is connectivity between people that promotes interactions and learning in urban spaces, between urban spaces, but also between urban and more remote locations. Connectivity is therefore central to the geography of the economy. However, the data and methodologies used to capture connectivity in economic geography have been limited due to the difficulty in gathering and analysing detailed observations in time and space about how people interact. Mobile phone data potentially offers a rich and unprecedented source of data, which is exhaustive in time and space and closely follows movements and partly communication activities of individuals. This paper provides a methodological overview of how mobile phone data has been used in studies related to economic geography, identifies the methodological challenges, elaborates on key findings for geography, and outlines opportunities for future research on the geography of connectivity using mobile phone data.

Since the late 20th 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 envisioned by Neil Smith (1993), with home, community and neighbourhoods, where interactions between individuals would occur on the streets and corners (Jacobs, 1969). Communication cuts across scales and allows for networks and structures to transcend distance with greater impact than before. It has prompted questions on the importance of spatial proximity, and critical reflections on whether this is the "death of geography" or the "death of distance" (Economist, 2003; Morgan, 2004). Whilst the speed of urbanisation and the continued concentration of knowledge intensive activities in space has debunked ideas about the death of distance (Iammarino et al., 2017; OECD, 2018; WorldBank, 2018), there is no doubt that the increased opportunities for communication have altered the geography of societies.

With the shift in use of mobile phones from mainly making calls into a devise that has become an essential part of everyday life, where individuals of most ages mostly always carry it with them, mobile phone data increasingly leaves a "digital trail" that closely follows individuals' movements and partly communication activities. Numerous methods have been developed for analysing mobile phone data and linking it to socio-economic factors. Studies using mobile phone data are scattered across various disciplines hardly noticed in the field of economic geography. While there are some reviews in the literature about the potential of big data for social sciences in general (e.g. Lazer and Radford, 2017: 20), a state of the art 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 through a systematic literature review (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 summarised 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.

II Methods

The literature review strictly follows transparent processes by systematically going through a cyclic process of obtaining, screening, and evaluating articles (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.

Table 1 approximately here

Second, the review uses a defined time frame. 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 pattern, and the correlation with socio-economic factors, it needs to have good coverage of the studied population. 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 mostly 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.

Ebscohost and Web of Science, two search engines covering partly overlapping and partly complementary academic databases, were used applying same search procedure with some exceptions: Ebscohost allowed for the Boolean search to be conducted on the entirety of the texts whilst Web of Science 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, presented in Table 2, 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 because it only spanned one database and had better options for limitations.

Table 2 approximately here

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 **Fehler! Verweisquelle konnte nicht gefunden werden.**. 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 (see Figure 1).

Figure 1 approximately here

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 Table 1. Studies of mobile phone data have contributed to three main themes: human mobility, social networks, and aggregate patterns of human mobility and social networks. Human mobility is a crosscutting theme, which often co-occurred with the other themes. Therefore, articles where only classified under this category if their main purpose was to investigate human mobility with mobile phone data, in total 69 articles. 22 studies focused on spatial differences of social networks. 28 articles investigated aggregate patterns of human mobility and social networks.

Figure 2 approximately here

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, 140 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 for economic geography, c) key findings, and d) limitations.

III How is mobile phone data used?

1 The nature of mobile phone data

The data generated by mobile phones has been called sensing data, sometimes also referred to as social sensing data. This mobile phone data is the output of the device when it interacts with its physical environment. This means that the dataset does not require direct or active intervention from the researcher in order to be generated. Sensing data includes any combination of a) location data, b) physical proximity to others, c) communication between users, d) users' interaction with the mobile phone, e) information stored in the device and lastly f) information about the device (Raento et al., 2009).

Location data is typically generated by its interaction with base receiver stations (BTS), also termed cell-towers (Frias-Martinez et al., 2012). This occurs when the phone either tries to contact other devices through calls or texts or when it uses the internet connection for various tasks (Lu et al., 2017; Rodriguez-Carrion et al., 2018). Location data can also be generated by using GPS data from the phone (Raento et al., 2009; Yadav et al., 2014). Physical proximity to others has often been observed through Bluetooth scans, which interact with nearby devices, allowing one to see which encounters a subject has over the duration of the study (Eagle et al., 2008; Raento et al., 2009). Communication is data generated when a user uses 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., 2009; Eagle et al., 2008; Raento et al., 2009). Information stored in the device, such as calendar information, and information about the device, e.g. if it is charging or on standby, requires a larger access to the individual devices and has in general been obtained through the use of certain applications (Eagle et al., 2009; Raento et al., 2009).

Most frequently used in the literature are call logs and location data, which often become available in what has been termed Call Details Records (CDR) (Lazer and Radford, 2017; Wei-Guang and Ming-Chia, 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 with 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 network that is perceived important instead of recording actual behaviour (Eagle et al., 2008). CDR are often geographically processed through its connection with the BTS's, which are geo-coded and cover a certain geographical area (Frias-Martinez et al., 2013; Hernandez et al., 2017; Moyano et al., 2012; Pappalardo et al., 2015; Vanhoof et al., 2018). There is also an alternative to the classical CDR, which has been termed mobile phone location data. It is similar to CDR but is not only generated when calls or text 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).

2 Major applications

Using sensing data, in particularly CDR, allows for two broad approaches which have been applied in research (Raento et al., 2009). One approach is to 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 other approach relies on sensing data only with the advantage of being able to cover a much larger sample. Location data is used to investigate how individuals (mobile phone devises) move in space and time. Using so-called Voronoi tessellations polygons, location data is geo-coded 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; Wang and Kilmartin, 2014; Vanhoof et al., 2018; 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. In addition, some studies using this approach have incorporated socio-economic data by aggregating Voronoi tessellations to fit the area of the socio-economic data (Cottineau and Vanhoof, 2019; Frias-Martinez et al., 2012; Pappalardo et al., 2015).

In both cases, the majority of studies are dealing with human mobility in some form. The popularity of using CDRs to capture human mobility 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 (see **table 7**). 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).

Table 3 approximately here

Methodologically, human mobility patterns are analysed with CDR using indices based on three types of measures: i) the travel distance, ii) the range of individuals' activity space, and iii) 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. It is primarily measured using the radius of gyration, which also has been used as an index for mobility volume (Gonzalez et al., 2008; Hoteit et al., 2014; Lu et al., 2017; Pappalardo et al., 2015). The radius of gyration, in essence, reflects the distance between a user's visited locations where a large value suggests a large range of activity space. 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 has been employed to study the similarities of mobility patterns. These range from relative simple ranking of indices (Becker et al., 2011), to more complex methods that uses 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 categorise 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; 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; Eagle et al., 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).

3 Merging mobile phone data with other 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. For economic data, Engelmann et al. (2018) used m-money transactions provided by the telecommunication provider 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. Another approach was done by Fixman et al. (2016) that 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 studied by Blumenstock (2018). In the end, these processes of merging supplementary data with CDR on individuals requires that both datasets can be joined, typically through the telephone ID.

On a regional level, there has been a number of different datasets that mobile phone data has been linked with. The 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 needed 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. (2010) and 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 organisation of cities.

4 Challenges and limitations

When handling mobile phone data there are a set of challenges and limitations that are generic across fields, which make it prone to bias and error (**table 4**). The limitations boil down to how the datasets are generated. The data is not primarily created for analytical purposes within scientific fields but for commercial purposes with the aim to collect information about customers and not about the population. If data from 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 because it will not be a random sampling of the population. However, mobile phone data can, despite its nature, represent sub-groups it has data from well, simply due to the large volume of users and usage within the dataset (Arai et al., 2016; Becker et al., 2011).

Table 4 approximately here

Another challenge relates 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 (see 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 (see 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 corresponds 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.

One challenge 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 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.

The challenge related to spatial and temporal scarcity can be addressed with 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. However, either solution do not deal with the challenges where an individual's presence in the dataset is due to its calling frequency, and thus the nature of its social network. Iovan et al. (2013) argued that it could create serious issues since the mobility patterns varies 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.

The issue of accuracy and completion also arises from how the data is generated when connecting to BTS (see e.g. Batran et al., 2018; Chen et al., 2018; Vanhoof et al., 2018). The BTSs are rarely placed equally across space and frequently follows existing urban structures with high concentrations in urban cores and low density in the rural peripheries. 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 (Batran 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.

As regards social network analysis, the challenges relate to the sheer amount of data in CDR datasets and to what this dataset actually reflects. 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).

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. De Montjoye et al. (2013) argued that it can be easy to identify individuals from anonymous CDR. This 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. 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; Frias-Martinez et al., 2013). 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). Having these caveats in mind, studies using mobile phone data have produced relevant findings about the geography of connectivity as discussed in the next section.

IV What are the key findings for human 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 i) human mobility, ii) social networks, and iii) aggregate patterns of human mobility and social networks.

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, 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; Järv et al., 2015).

Studies on individual mobility identify regular patterns that hold across cultures and spatial contexts (Gonzalez et al., 2008; Kang et al., 2012; Yuan and Raubal, 2014). In particular, the activity space declines with distance. The distance decay has been modelled with a power law distribution (Gonzalez et al., 2008; Lixing et al., 2017; Moyano et al., 2012) or an exponential law distribution for intra-urban travel patterns (Kang et al., 2012). In line with this, the radius of gyration follows the same pattern

(Gonzalez et al., 2008). This regularity implies that individuals' activity spaces can be captured within a short period of time and consequently predicted (Song et al., 2010).

However, despite general regularities, individual's 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. A similar observation was made by Yadav et al. (2014) who 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 regular patterns of individual mobility have been used to estimate 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). 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 (Pappalardo and Simini, 2018).

2 Individuals social networks

Findings on individuals' social networks from mobile phone data relate to three categories: i) the connection between social networks and human mobility, ii) the differences between spatial contexts, and iii) 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 itself 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 amount 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 amount 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.

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 et al., 2015; Blumenstock and Eagle, 2010; Blumenstock, 2018). 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. (2010) & 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. For instance, Cottineau and Vanhoof (2019) found that mobility range and diversity tended to decrease in cities with large levels of segregation. Järv 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.

V Concluding discussion

Mobile phone service providers worldwide have access to data from almost 8 billion mobile phone subscribers, a number that almost doubled over the last 10 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.

Mobile phone data is 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 methods 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" but focuses 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 etc. 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 human geography in general and economic geography in specific, a major drawback is the lack of theorizing (Pappalardo et al., 2015). The majority of studies has been conducted by computer scientists who are not necessarily trained in geographic thought (Lazer and Radford, 2017). As a consequence, research using mobile phone data has the tendency of being descriptive and of mainly focussing on correlations with a conceivable lack of theoretical connections.

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 provides 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 is 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 mobilised 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. This data is 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 has a high resolution in time and in scale 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 increases the potential for causal inference.

Adding data on people's location in 5-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 socioeconomic 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 is 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. 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.

References

Ahas R, Aasa A, Yuan Y, et al. (2015) Everyday space-time geographies: using mobile phone-based sensor data to monitor urban activity in Harbin, Paris, and Tallinn. *International Journal of Geographical Information Science* 29: 2017-2039.

Amini A, Kung K, Kang CG, et al. (2014) The impact of social segregation on human mobility in developing and industrialized regions. *Epj Data Science* 3: 20.

Andersson M and Karlsson C. (2007) Knowledge in Regional Economic Growth—The Role of Knowledge Accessibility. *Industry and Innovation* 14: 129-149.

Arai A, Fan ZP, Matekenya D, et al. (2016) Comparative Perspective of Human Behavior Patterns to Uncover Ownership Bias among Mobile Phone Users. *Isprs International Journal of Geo-Information* 5: 85.

Bajardi P, Delfino M, Panisson A, et al. (2015) Unveiling patterns of international communities in a global city using mobile phone data. *Epj Data Science* 4: 17.

Barbosa H, Barthelemy M, Ghoshal G, et al. (2018) Human mobility: Models and applications. *Physics Reports* 734: 1-74.

Batran M, Mejia MG, Sekimoto Y, et al. (2018) Inferencing Human Spatiotemporal Mobility in Greater Maputo via Mobile Phone Big Data Mining. *Isprs International Journal of Geo-Information* 7: 259.

Becker RA, Caceres R, Hanson K, et al. (2011) A Tale of One City: Using Cellular Network Data for Urban Planning. *Ieee Pervasive Computing* 10: 18-26.

Blumenstock J, Cadamuro G and On R. (2015) Predicting poverty and wealth from mobile phone metadata. *Science* 350: 1073-1076.

Blumenstock J and Eagle N. (2010) Mobile divides: gender, socioeconomic status, and mobile phone use in Rwanda. *Proceedings of the 4th ACM/IEEE International Conference on Information and Communication Technologies and Development.* ACM, 6.

Blumenstock JE. (2018) Estimating Economic Characteristics with Phone Data. *AEA Papers and Proceedings* 108: 72-76.

Calabrese F, Diao M, Di Lorenzo G, et al. (2013) Understanding individual mobility patterns from urban sensing data: A mobile phone trace example. *Transportation Research: Part C* 26: 301-313.

Calabrese F, Smoreda Z, Blondel VD, et al. (2011) Interplay between telecommunications and face-to-face interactions: A study using mobile phone data. *PLoS ONE* 6: e20814.

Castillo G, Layedra F, Guaranda MB, et al. (2018) The Silence of the Cantons: Estimating Villages Socioeconomic Status Through Mobile Phones Data. *2018 International Conference on eDemocracy & eGovernment (ICEDEG).* Place of Publication: Piscataway, NJ, USA; Ambato, Ecuador. Country of Publication: USA.: IEEE.

Chen GS, Hoteit S, Viana AC, et al. (2018) Enriching sparse mobility information in Call Detail Records. *Computer Communications* 122: 44-58.

Chi G, Thill JC, Tong D, et al. (2016) Uncovering regional characteristics from mobile phone data: A network science approach. *Papers in Regional Science* 95: 613-631.

Cortinovis N, Xiao J, Boschma R, et al. (2017) Quality of government and social capital as drivers of regional diversification in Europe. *Journal of Economic Geography* 17: 1179-1208.

Cottineau C and Vanhoof M. (2019) Mobile Phone Indicators and Their Relation to the Socioeconomic Organisation of Cities. *Isprs International Journal of Geo-Information* 8: 19.

Dash M, Koo KK, Holleczek T, et al. (2015) From Mobile Phone Data to Transport Network - Gaining Insight About Human Mobility. *2015 16th IEEE International Conference on Mobile Data Management, Vol 1.* New York: Ieee, 243-250.

Dashdorj Z, Sobolevsky S, Serafini L, et al. (2014) Human activity recognition from spatial data sources. In: Chow CY and Shekhar S (eds) *Proceedings of the Third Acm Sigspatial International Workshop on Mobile Geographic Information Systems.* New York: Assoc Computing Machinery, 18-25.

De Montjoye Y-A, Hidalgo CA, Verleysen M, et al. (2013) Unique in the crowd: The privacy bounds of human mobility. *Scientific Reports* 3: 1376.

De Montjoye Y-A, Rocher L and Pentland AS. (2016) bandicoot: A python toolbox for mobile phone metadata. *The Journal of Machine Learning Research* 17: 6100-6104.

Doyle C, Herga Z, Dipple S, et al. (2019) Predicting complex user behavior from CDR based social networks. *Information Sciences* 500: 217-228.

Eagle N, de Montjoye YA and Bettencourt LMA. (2009) Community computing: comparisons between rural and urban societies using mobile phone data. *2009 International Conference on Computational Science and Engineering (CSE).* Place of Publication: Piscataway, NJ, USA; Vancouver, BC, Canada. Country of Publication: USA.: IEEE.

Eagle N, Macy M and Claxton R. (2010) Network diversity and economic development. *Science* 328: 1029–1031.

Eagle N, Pentland A and Lazer D. (2008) *Mobile phone data for inferring social network structure,* New York: Springer.

Economist. (2003) The revenge of geography. *The Economist.* S19-S22.

Engelmann G, Smith G and Goulding J. (2018) The unbanked and poverty: predicting area-level socio-economic vulnerability from m-money transactions. *2018 IEEE International Conference on Big Data (Big Data).* Piscataway, NJ, USA; Seattle, WA, USA. : IEEE.

Ettlinger N. (2003) Cultural economic geography and a relational and microspace approach to trusts, rationalities, networks, and change in collaborative workplaces. *Journal of Economic Geography* 3: 145-171.

Fixman M, Berenstein A, Brea J, et al. (2016) A Bayesian approach to income inference in a communication network. *2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM).* Place of Publication: Piscataway, NJ, USA; San Francisco, CA, USA. Country of Publication: USA.: IEEE.

Frias-Martinez V, Soto V, Virseda J, et al. (2013) Can cell phone traces measure social development. *Third Conference on the Analysis of Mobile Phone Datasets, NetMob.*

Frias-Martinez V, Virseda-Jerez J and Frias-Martinez E. (2012) On the relation between socio-economic status and physical mobility. *Information Technology for Development* 18: 91-106.

Frias-Martinez V, Virseda J and Frias-Martinez E. (2010) Socio-economic levels and human mobility. *Qual meets quant workshop-QMQ.*

Gaito S, Quadri C, Rossi GP, et al. (2017) Urban communications and social interactions through the lens of mobile phone data. *Online Social Networks and Media* 1: 70-81.

Ghosh I and Singh VK. (2018) Modeling social support scores using phone use patterns. *Proceedings of the Association for Information Science and Technology* 55.

Giuliani E. (2007) The selective nature of knowledge networks in clusters: evidence from the wine industry. *Journal of Economic Geography* 7: 139-168.

Gonzalez MC, Hidalgo CA and Barabasi A-L. (2008) Understanding individual human mobility patterns. *Nature* 453: 779.

Grillitsch M and Nilsson M. (2015) Innovation in peripheral regions: Do collaborations compensate for a lack of local knowledge spillovers? *The Annals of Regional Science* 54: 299-321.

Hall O, Hay GJ, Bouchard A, et al. (2004) Detecting dominant landscape objects through multiple scales: An integration of object-specific methods and watershed segmentation. *Landscape Ecology* 19: 59-76.

Hernandez M, Hong L, Frias-Martinez V, et al. (2017) Estimating poverty using cell phone data: evidence from Guatemala. *Policy Research Working Paper Series.* The World Bank.

Hoteit S, Secci S, Sobolevsky S, et al. (2014) Estimating human trajectories and hotspots through mobile phone data. *Computer Networks* 64: 296-307.

Hägerstrand T. (1970) What about people in regional science? *Papers in Regional Science* 24: 7-24.

Iammarino S, Rodríguez-Pose A and Storper M. (2017) Why Regional Development matters for Europe's Economic Future. *Working Papers of the Directorate-General for Regional and Urban Policy*.

Iovan C, Olteanu-Raimond AM, Couronne T, et al. (2013) Moving and Calling: Mobile Phone Data Quality Measurements and Spatiotemporal Uncertainty in Human Mobility Studies. In: Vandenbroucke D, Bucher B and Crompvoets J (eds) *Geographic Information Science at the Heart of Europe.* Berlin: Springer-Verlag Berlin, 247-265.

Jacobs J. (1969) *The Economy of Cities,* New York: Random House.

Jahani E, Sundsøy P, Bjelland J, et al. (2017) Improving official statistics in emerging markets using machine learning and mobile phone data. *Epj Data Science* 6: 3.

Järv O, Ahas R and Witlox F. (2014) Understanding monthly variability in human activity spaces: A twelve-month study using mobile phone call detail records. *Transportation Research Part C: Emerging Technologies* 38: 122-135.
Järv O, Müürisepp K, Ahas R, et al. (2015) Ethnic differences in activity spaces as a characteristic of segregation: A study based on mobile phone usage in Tallinn, Estonia. *Urban Studies* 52: 2680-2698.
Kang C, Ma X, Tong D, et al. (2012) Intra-urban human mobility patterns: An urban morphology perspective. *Physica A* 391: 1702-1717.
Karikoski J and Nelimarkka M. (2011) Measuring social relations with multiple datasets. *International Journal of Social Computing and Cyber-Physical Systems* 1: 98-113.
Kemeny T, Feldman M, Ethridge F, et al. (2016) The economic value of local social networks. *Journal of Economic Geography* 16: 1101-1122.
Kung KS, Greco K, Sobolevsky S, et al. (2014) Exploring Universal Patterns in Human Home-Work Commuting from Mobile Phone Data. *PLoS ONE* 9: 1-15.
Lazer D and Radford J. (2017) Data ex Machina: Introduction to Big Data. In: Cook KS and Massey DS (eds) *Annual Review of Sociology, Vol 43.* Palo Alto: Annual Reviews, 19-39.
Li M, Gao S, Lu F, et al. (2019) Reconstruction of human movement trajectories from large-scale low-frequency mobile phone data. *Computers, Environment & Urban Systems* 77: N.PAG-N.PAG.
Lixing S, Wen W, Wei C, et al. (2017) Mobility patterns analysis of Beijing residents based on call detail records. *2017 9th International Conference on Wireless Communications and Signal Processing (WCSP).* Place of Publication: Piscataway, NJ, USA; Nanjing, China. Country of Publication: USA.: IEEE.
Lu SW, Fang ZX, Zhang XR, et al. (2017) Understanding the Representativeness of Mobile Phone Location Data in Characterizing Human Mobility Indicators. *Isprs International Journal of Geo-Information* 6: 19.
Mamei M, Pancotto F, De Nadai M, et al. (2018) Is social capital associated with synchronization in human communication? An analysis of Italian call records and measures of civic engagement. *Epj Data Science* 7: 16.
Manfredini F, Pucci P and Tagliolato P. (2013) Deriving Mobility Practices and Patterns from Mobile Phone Data. In: Murgante B, Misra S, Carlini M, et al. (eds) *International Conference on Computational Science and Its Applications.* Berlin: Springer-Verlag Berlin, 438-451.
Mao HN, Shuai X, Ahn YY, et al. (2015) Quantifying socio-economic indicators in developing countries from mobile phone communication data: applications to Cote d'Ivoire. *Epj Data Science* 4: 16.
Morgan K. (2004) The exaggerated death of geography: learning, proximity and territorial innovation systems. *Journal of Economic Geography* 4: 3-21.
Mota T, Munjal A and Camp T. (2015) Large-scale human mobility analysis based on mobile phone and social media communication: a case-study in Africa. *2015 16th IEEE International Conference on Mobile Data Management (MDM).* Place of Publication: Los Alamitos, CA, USA; Pittsburgh, PA, USA. Country of Publication: USA.: IEEE Computer Society.
Moyano LG, Thomae ORM and Frias-Martinez E. (2012) Uncovering the Spatio-Temporal Structure of Social Networks using Cell Phone Records. In: Vreeken J, Ling C, Zaki MJ, et al. (eds) *12th IEEE International Conference on Data Mining Workshops.* New York: Ieee, 242-249.
OECD. (2018) *Productivity and Jobs in a Globalised World.*
Onnela J-P, Arbesman S, González MC, et al. (2011) Geographic constraints on social network groups. *PLoS ONE* 6: e16939.
Pappalardo L and Simini F. (2018) Data-driven generation of spatio-temporal routines in human mobility. *Data Mining and Knowledge Discovery* 32: 787-829.
Pappalardo L, Smoreda Z, Pedreschi D, et al. (2015) Using Big Data to study the link between human mobility and socio-economic development. *2015 IEEE International Conference on Big Data (Big Data).* Place of Publication: Piscataway, NJ, USA; Santa Clara, CA, USA. Country of Publication: USA.: IEEE.
Phithakkitnukoon S, Smoreda Z and Olivier P. (2012) Socio-Geography of Human Mobility: A Study Using Longitudinal Mobile Phone Data. *PLoS ONE* 7: 1-9.
Puura A, Silm S and Ahas R. (2018) The Relationship between Social Networks and Spatial Mobility: A Mobile-Phone-Based Study in Estonia. *Journal of Urban Technology* 25: 7-25.

Raento M, Oulasvirta A and Eagle N. (2009) Smartphones: An emerging tool for social scientists. *Sociological methods & research* 37: 426-454.

Rodriguez-Carrion A, Garcia-Rubio C and Campo C. (2018) Detecting and Reducing Biases in Cellular-Based Mobility Data Sets. *Entropy* 20: 736.

Rodríguez-Pose A. (2013) Do Institutions Matter for Regional Development? *Regional Studies* 47: 1034-1047.

Šćepanović S, Mishkovski I, Hui P, et al. (2015) Mobile Phone Call Data as a Regional Socio-Economic Proxy Indicator. *PLoS ONE* 10: 1-15.

Schmid T, Bruckschen F, Salvati N, et al. (2017) Constructing sociodemographic indicators for national statistical institutes by using mobile phone data: estimating literacy rates in Senegal. *Journal of the Royal Statistical Society Series a-Statistics in Society* 180: 1163-1190.

Silm S, Ahas R and Mooses V. (2018) Are younger age groups less segregated? Measuring ethnic segregation in activity spaces using mobile phone data. *Journal of Ethnic & Migration Studies* 44: 1797-1817.

Simini F, González MC, Maritan A, et al. (2012) A universal model for mobility and migration patterns. *Nature* 484: 96.

Singh VK and Ghosh I. (2017) Inferring Individual Social Capital Automatically via Phone Logs. *Proceedings of the ACM on Human-Computer Interaction* 1: 95.

Smith N. (1993) Homeless/global: scaling places. In: Bird J, Curtis B, Putnam T, et al. (eds) *Mapping the futures: local cultures, global change.* London: Routledge.

Song C, Qu Z, Blumm N, et al. (2010) Limits of predictability in human mobility. *Science* 327: 1018-1021.

Thuillier E, Moalic L, Lamrous S, et al. (2018) Clustering Weekly Patterns of Human Mobility Through Mobile Phone Data. *Ieee Transactions on Mobile Computing* 17: 817-830.

Toole JL, Lin Y-R, Muehlegger E, et al. (2015) Tracking employment shocks using mobile phone data. *Journal of The Royal Society Interface* 12: 20150185.

Tranfield D, Denyer D and Smart P. (2003) Towards a methodology for developing evidence-informed management knowledge by means of systematic review. *British journal of management* 14: 207-222.

Wang H and Kilmartin L. (2014) Comparing Rural and Urban Social and Economic Behavior in Uganda: Insights from Mobile Voice Service Usage. *Journal of Urban Technology* 21: 61-89.

Vanhoof M, Schoors W, Van Rompaey A, et al. (2018) Comparing Regional Patterns of Individual Movement Using Corrected Mobility Entropy. *Journal of Urban Technology* 25: 27-61.

Wei-Guang T and Ming-Chia C. (2007) Mining communities of acquainted mobile users on call detail records. *Applied Computing 2007. The 22nd Annual ACM Symposium on Applied Computing.* Place of Publication: New York, NY, USA; Seoul, South Korea. Country of Publication: USA.: ACM.

WorldBank. (2018) Growing united: upgrading Europe's convergence machine. Washington, D.C.: World Bank Group.

Yadav K, Kumar A, Bharati A, et al. (2014) Characterizing mobility patterns of people in developing countries using their mobile phone data. *2014 Sixth International Conference on Communication Systems and Networks (COMSNETS).* Place of Publication: Piscataway, NJ, USA; Bangalore, India. Country of Publication: USA.: IEEE.

Yang XP, Fang ZX, Xu Y, et al. (2016) Understanding Spatiotemporal Patterns of Human Convergence and Divergence Using Mobile Phone Location Data. *Isprs International Journal of Geo-Information* 5: 18.

Yuan YH and Raubal M. (2014) Measuring similarity of mobile phone user trajectories - a Spatio-temporal Edit Distance method. *International Journal of Geographical Information Science* 28: 496-520.

Yuan YH and Raubal M. (2016) Analyzing the distribution of human activity space from mobile phone usage: an individual and urban-oriented study. *International Journal of Geographical Information Science* 30: 1594-1621.

Zhao ZL, Shaw SL, Xu Y, et al. (2016) Understanding the bias of call detail records in human mobility research. *International Journal of Geographical Information Science* 30: 1738-1762.

Table 5: Table of Keywords used to collect relevant articles

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

Table 6: Academic databases and subject areas

	Ebscohost	*Web of Science*
Databases	Academic Search Complete Bibliography of Asian Studies Communication Source eBook Collection EconLit GreenFile Humanities international complete Inspec MathSciNet SocINDEX Urban Studies Abstracts	Web of Science Core Collection
Subject Areas	NA	Geosciences, Multidisciplinary Social Sciences, Mathematical Methods Mathematics Mathematics, Applied Sociology Computer Science, Information Systems Remote Sensing Urban Studies

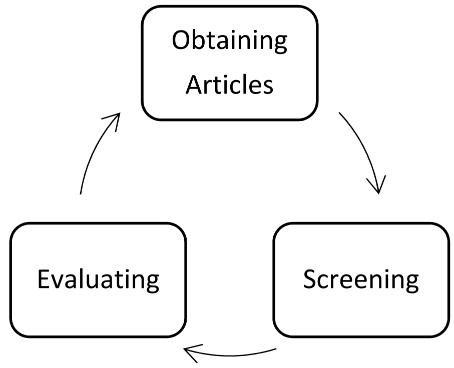
Table 7: Key variables for the analysis of location and call data (Pappalardo et al. 2015)

Variable	Caller (phone)	Callee (phone)	Timestamp	Telephone Mast	Tower's coordinate
Format	Unique ID	Unique ID	Year/month/day time	Unique ID	Latitude/Longitude
Example	999J00	000J99	2019/11/06 10:57	5	55.70886, 13.200803

Table 8: Limitations of mobile phone data

	Limitations	*Key issues*	*Main Consequences*	*Suggested approaches*
Data	Bias	• Ownership bias(Arai et al. 2016): o Age o Income o Gender • Individual variance in calling patterns (Puura, Silm, and Ahas 2018) • Space specific calling patterns (Zhao et al. 2016)	• Questions the representability for larger communities and regions • Variance inflation in predictions • Skews the human mobility pattern to emphasise specific places	• Strive for records that has greater coverage of the population to minimise ownership bias • If available, detailed records contain far more detailed trails of mobility and communication. Limiting the influence of space specific patterns (Cottineau and Vanhoof 2019; Rodriguez-Carrion, Garcia-Rubio, and Campo 2018)
	Source	• Commercially generated data (Lazer and Radford 2017) • Generated only when used (Chen et al. 2018)	• Cannot achieve random sampling of the population (Iovan et al. 2013) • Usage of the phone does not necessarily align with travel behaviour (Zhao et al. 2016)	• Aggregating mobility patterns to weekly patterns (Thuillier et al. 2018) • Increasing the observation period (Batran et al. 2018)
	Quality	• Overlapping BTS radius o Nearest tower is not necessarily picked (Rodriguez-Carrion, Garcia-Rubio, and Campo 2018) • Varying detail of mobility within the datasets (Vanhoof et al. 2018) o Spatial difference in BTS density	• Variance inflation in mobility and calling patterns (Cottineau and Vanhoof 2019; Vanhoof et al. 2018)	• Weighting the dataset to account for the spatial difference of BTS density and its effect on mobility and communication values (Vanhoof et al. 2018)
Processing	Time	• Large datasets in need of multiple processing steps (De Montjoye, Rocher, and Pentland 2016) o Privacy concerns (De Montjoye et al. 2013) o Data cleaning	• Time consuming processes that can prove costly for research (Frias-Martinez et al. 2013) • Increasing difficulties in terms of transparency	• Streamline the data cleaning and analysis by using the python library "bandicoot" (De Montjoye, Rocher, and Pentland 2016)
	Accuracy	• Data cleaning and calculation moves data further away from the ground truth (Zhao et al. 2016; Lazer and Radford 2017; Iovan et al. 2013) • Aggregation of data to appropriate level increasingly deals with MAUP (Cottineau and Vanhoof 2019)	• Caution needs to be applied to the strengths of the conclusions that can be made from the analysis	• Increase granularity when possible and analyse the impact of scales on the results
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 (Pappalardo et al. 2015; Lazer and Radford 2017)	• Tendency to descriptive correlation analysis in socio-economic connections.	•

Figure 2: The process of obtaining a foundation of articles for the literature review



	Boolean search process		Referenced articles
	Ebscohost	Web of Science	
Obtained articles	427	189	30
Screened articles	98	72	30
Evaluated articles	61	53	27

Figure 2: Papers by topic and type of source