

Article

# Are the Poor Digitally Left Behind? Indications of Urban Divides Based on Remote Sensing and Twitter Data

Hannes Taubenböck <sup>1,\*</sup> , Jeroen Staab <sup>1</sup> , Xiao Xiang Zhu <sup>1,2</sup> , Christian Geiß <sup>1</sup>, Stefan Dech <sup>1</sup> and Michael Wurm <sup>1</sup> 

<sup>1</sup> German Aerospace Center (DLR), German Remote Sensing Data Center (DFD), Münchner Straße 20, 82234 Weßling, Germany; jeroen.staab@dlr.de (J.S.); xiao.zhu@dlr.de (X.X.Z.); christian.geiss@dlr.de (C.G.); stefan.dech@dlr.de (S.D.); michael.wurm@dlr.de (M.W.)

<sup>2</sup> Technical University of Munich (TUM), Signal Processing in Earth Observation (SiPEO), 80333 Munich, Germany

\* Correspondence: hannes.taubenboeck@dlr.de

Received: 20 June 2018; Accepted: 25 July 2018; Published: 30 July 2018



**Abstract:** Every city is—quoting Plato—divided into two, one city of the poor, the other of the rich. In this study we test whether the economic urban divide is reflected in the digital sphere of cities. Because, especially in dynamically growing cities, ready-to-use comprehensive data sets on the urban poor, as well as on the digital divide, are not existent, we use proxies: we spatially delimit the urban poor using settlement characteristics derived from remote sensing data. The digital divide is targeted by geolocated Twitter data. Based on a sample of eight cities across the globe, we spatially test whether areas of the urban poor are more likely to be digital cold spots. Over the course of time, we analyze whether temporal signatures in poor urban areas differ from formal environments. We find that the economic divide influences digital participation in public life. Less residents of morphological slums are found to be digitally oriented (“are digitally left behind”) as compared to residents of formal settlements. However, among the few twitter users in morphological slums, we find their temporal behavior similar to the twitter users in formal settlements. In general, we conclude this discussion, this study exemplifies that the combination of both heterogeneous data sets allows for extending the capabilities of individual disciplines for research towards urban poverty.

**Keywords:** urban society; morphological slum; digital divide; social media; Twitter; remote sensing; geoinformation system; human geography

## 1. Introduction

“Every city, however small, is, in fact, divided into two, one city of the poor, the other of the rich; [ . . . ]” [1] after Plato. With 32.7% of global city dwellers estimated to live in slums [2], with an expected global population growth of 2.3 billion people until 2050 [3], and in a world where place of birth as well as living environments have major influence on (economic) wealth [4], the pressure on cities and their societies is increasing. Challenges related to the process of urbanization are echoed by intergovernmental agreements on the Sustainable Development Goals (SDGs) [3]: ‘To end poverty’ (SDG1) and ‘to build sustainable cities and communities’ (SDG 11) are, among others, development goals explicitly related to urban poverty. For creating sustainable development strategies, data and information are crucial. However, for the ‘city of the poor’ most countries lack adequate data to identify and monitor these places, to understand processes, their inhabitants, their behavior, etc. Although we live in an era in which more (geo)data are available than ever before in human history, the World Migration Report [5] states “we face a massive lack of basic data about urban poverty”; thus, the poorest

people often remain invisible in statistics [5]. And, if data are available, the credibility of these data is in doubt [6,7].

In this paper, we search for new ways of big data analytics to reduce knowledge gaps on urban poverty, i.e., by the multidisciplinary combination of different data types from *remote sensing* and *social networks*. Along the argument of [8], suggesting to go “*beyond the geotag*” by leveraging data from social networks against ancillary data, we investigate whether these two different data types can be complementarily used to exploit social phenomena regarding the urban divide. In this study, we understand the term ‘*urban divide*’ by two different aspects: The *economic divide* characterizing different social groups within the city. The *digital divide* exhibits different participations in online resources by different social groups.

The initial quote in this paper provocatively culminates a simplistic, dichotomic perspective towards the urban divide. Naturally, the ‘*city of the poor*’ and the ‘*city of the rich*’ are neither geographically nor socio-economically strictly separated from each other, but they are interwoven in complex, temporarily dynamic patterns of living, working, traveling, communicating, and many other aspects. However, due to the lack of adequate data in many cities on these multidimensional interwoven complex patterns, we approach the urban divide by applying commonly used proxies. We investigate whether residential places of the urban poor, mapped by remote sensing, show different spatio-temporal patterns of social network activity. Methodically, we quantitatively compare the number of tweets across space and, we investigate tweet frequencies over the course of time. By doing so, we aim to identify trends in human behavior at specific urban locations (i.e., morphological slums) and whether these trends are similar. For investigating this social phenomenon, we apply this approach at multiple cities across the globe.

The remainder of the work is structured as follows: Section 2 introduces the background on urban poverty and presents the rationale of this work. Section 3 introduces the experimental set-up with the selected cities under investigation, the data sets used, the methodology of mapping urban poverty using Earth observation (EO)-data, the pre-processing of social media data and the combined analyses of both data sets. In Section 4 the results on spatio-temporal social media activity for varying living environments within the city are presented. This is followed by a discussion in Section 5, where capabilities and limitations of the data sets are evaluated. Section 6 concludes with a perspective.

## 2. Background and Rationale

The scientific debate on urban poverty comprises a broad range of topics. These include measurements of poverty either in an economic, e.g., [9], a spatial, e.g., [10], or a multidimensional, e.g., [11] sense. Processes leading to urban poverty are analyzed, e.g., [12,13], questions on why the poor prefer to live in cities are investigated, e.g., [14], physical living environments of the urban poor are characterized and localized, e.g., [15], social phenomena, e.g., [16] such as effects of slum-upgrading on social cohesion are described [17], among many other studies about the urban poor. From a political perspective, the multitudes of challenges for the urban poor (and the global society as a whole) are on the political agenda (e.g., SDGs) [3]. However, the demand for ‘sustainable data for sustainable development’, i.e., improved data availability, quality, consistency, timelines and disaggregation is often not met and creates a demand for exploring new (multidisciplinary) data sets and methods.

In this study, we explore the capabilities of combining two data sets—*remote sensing* and *social network* data. We investigate urban poverty by addressing the *economic divide* as well as the *digital divide* within cities. We understand the term ‘urban divide’ as two social groups’ representative for the differentiation between the marginalized and poor urban residents on the one hand, and formal and economically better off residents on the other.

Addressing the *economic divide* is complex as the measurement of poverty, and it is not unambiguous. Varying methods such as conventional economic vs. participatory anthropological approaches have been discussed, e.g., [18]. Commonly poverty is defined at household level; relative measures usually relate to households having less than 50% of the median income of the population

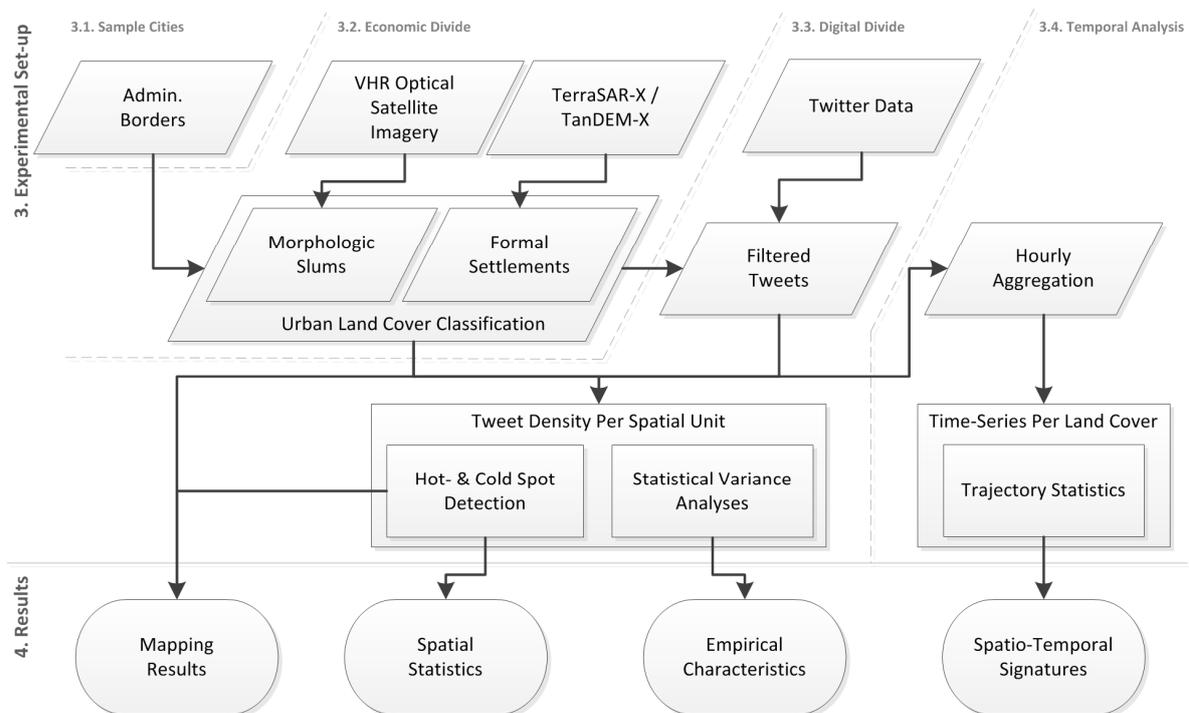
within a specific geographic entity [9]. However, these or related socio-economic data for the urban poor are rarely available (or accessible) at household level in many cities (especially in the Global South). When available, these statistics are mostly aggregated, outdated, or inconsistent [5,19]. Against this background, in this study we approach the economic divide spatially by the delimitation of neighborhoods with poor building structures. Mapping places of residence of the urban poor using *remote sensing* data has gained increasing attention by scholars since very high resolution (VHR) satellite data became available in 1999, see review by [19]. Characteristic structures of the built environment (organic lay-outs, high densities, small ground floors, among others) are generally used as proxy for delimiting living areas of slums [10,15,20]. Manifold image classification approaches allow the spatial classification of these places from VHR optical, e.g., [21,22] and radar, e.g., [23] data. Nevertheless, this proxy also features shortcomings, as the inhabitants in these areas are naturally not a fully homogeneous social group, nor is it vice versa. Although an inherent binary classification into 'poor' and 'not poor' neglects the complexity of the spatial distribution of social groups in space, studies prove that these spatially delimited areas are a feasible and legitimate proxy for localizing living environments of the social group of urban poor [24,25]. Beyond, using characteristic structures of the built environment allows for spatial consistency within and across cities, and at least with its shortcomings, it provides completeness of comparable spatial extents. With it, we intrinsically follow a spatial approach localizing places of residence.

The *digital divide* is commonly understood as a multidimensional phenomenon encompassing divergence of Internet access across space (global divide), the gap between information rich and poor areas (social divide), as well as the gap within online users, i.e., between those who do, and do not, use the panoply of digital resources to participate in the public life (democratic divide) [26]. Ref. [27] finds, that most frequently the literature on the digital divide identifies the poor as being most vulnerable to negative impacts from participating in the online world. In this regard, *social network data* have recently gained increasing attention by scholars to capture, analyze and understand the digital divide, e.g., [28–30]. Focusing on urban poverty investigations reveal the relevance of social networks for empowering people to participate in urban societies [31,32]. However, it is also shown that urban inequalities with respect to usage of such communication platforms remain [33]. Other investigations apply social network data for capturing dynamic patterns and human interactions. For instance, scholars map urban land use by spatio-temporal activity patterns, e.g., [34], unusual events, e.g., [35], mobility patterns, e.g., [36] or they define and characterize neighborhoods, e.g., [37]. Against this background, we approach the digital divide in this study by data from the global microblogging platform Twitter. The location-based twitter data functions as a proxy for the digital landscape of a city. Nevertheless, we are aware that the twitter users cannot be taken as representative of the entire population of the users being online. The proxy contains a highly non-uniform sample of the entire population with inherent biases, e.g., [38,39]. Ref. [40], as an example, find a strong bias towards male users; ref. [30] as well as [41] find a bias towards young, highly educated males with social upward mobility using these communication streams versus lower-income residents tending to be less digitally oriented. Focusing on urban poor living in slum environments, we suppose that technical aspects, such as access to electricity or to devices with online access, as well as social aspects such as literacy rate or education levels, influence the participation in these social online platforms. Further, it is important to note, that residents of morphological slums can also tweet from within formal settlement structures influencing our statistics. However, despite these restrictions, we assume the site-specific tweet activity and the respective temporal patterns allow the general characterization of behaviors of residents and presence of infrastructures related to modern communication technologies.

As a result, we combine the two data sets from *satellite imagery* and *twitter* to approach the following specific research questions: (1) Are the physical built-up structures, used as proxy for the economic divide in cities, reflecting the digital divide within the population of twitter users? (2) And, do locations of urban poor feature different temporal behavioral patterns of twitter activity than formal settlements?

### 3. Experimental Set-Up: Materials and Methods

In general, we follow in our experimental set-up a hierarchical rationale by first mapping the *economic divide*, and secondly, investigating whether spatially related behavioral patterns in social networks participation exist and confirm a *digital divide*. The workflow is illustrated in Figure 1: *First*, a sample of cities for our investigation is selected. *Second*, using the multi-source remote sensing data, the settlement areas of our selected experimental cities are spatially sub-divided into the thematic classes ‘morphologic slums’ and ‘formal settlements’. *Third*, the twitter data are preprocessed and filtered to produce an unbiased sample. *Fourth*, two different types of analyses are conducted, which are illustrated in Figure 1 as rectangles: One type uses the spatial tweet densities and the other using spatio-temporal frequencies. The spatial tweet density analysis relies on the number of tweets per spatial unit; thus, both datasets are aggregated on spatial units used within the respective administrative area of the studied cities. With it, digital hot and cold spots are detected and statistical variance analyses are conducted. The results are presented as mapping representation, by empirical characteristics at cross-city and intra-urban scales and spatial statistics. The spatio-temporal analysis relies on the number of tweets per hour and day. The correspondingly aggregated temporal signatures per land cover class outline user behavior of different social groups.



**Figure 1.** Workflow illustrating the experimental set-up, i.e., data, methodological steps of analysis. Beyond, the varying results are presented. All steps are explained in detail below with the chapter numberings indicating the outline in the paper.

#### 3.1. Selection of Experimental Cities

For the experiments we select a meaningful sample of study sites by the following reason:

- We select cities which feature a significant share of urban poor documented in literature or official census data.
- We select cities containing building morphologies characteristic for slums, which are in line with the conceptual ontology presented by [20] and the related empirical basis demonstrated by [15].

This physical appearance differs to formal settlements and can be classified by VHR optical remote sensing data.

- We select cities at different cultural areas and continents across the globe.

In addition to these attributes our selection process is guided by the limitation that every chosen city needs an extensive and consistent classification of morphologic slum structures based on available VHR optical remote sensing data. As of today, no such classification is generally or globally available and as it has been reported in a comprehensive literature review [19], only few case studies exist. In consequence, we limited the selection to eight cities, where consistent classifications of morphologic slums in line with our conceptual ontology have been produced (cf. Section 3.2.2). Based on these criteria, our selected sample contains heterogeneous cities by e.g., city size, population, economy, cultural situation or politics: *Dhaka* (Bangladesh), *Mumbai* (India), *Manila* (Philippines) in Asia; *Caracas* (Venezuela) and *Rio de Janeiro* (Brazil) in America; *Cairo* (Egypt) and *Cape Town* (South Africa) in Africa; and *Lisbon* (Portugal) in Europe.

Our *general area under investigation* is the administrative area for each city. Therefore, we use the spatial outlines provided by the Global Administrative Areas (GADM) database [42].

### 3.2. Mapping the Economic Divide

#### 3.2.1. Remote Sensing Data

For the localization of living environments of different social groups within the urban landscape *remote sensing data* are applied. For the analysis multi-source earth observation data from VHR optical satellite sensor systems (1) as well as from high resolution radar systems (2) are used:

1. Optical sensors, such as QuickBird or WorldView, provide geometric resolutions of 1m and better and thus, the urban morphology is represented by individual buildings. We apply these data for the delimitation of morphologic slums. Figure 2a illustrates the complex urban environment by contrasting geometric, planned, formal building structures with non-regular, unplanned, non-formal building structures of morphological slums.
2. We use radar data from the TerraSAR-X and TanDEM-X missions at Stripmap mode providing geometric resolutions of 3 m. For urban landscapes spatial complexity of varying objects within small areas is characteristic. In radar data this is represented in highly textured image regions of strong directional, non-Gaussian backscatter due to double bounce effects. This information is used along with the intensity information to delineate 'settlements' from 'non-settlements' using an unsupervised image analysis technique, for technical details we refer to [43]. The accuracies of the settlement classification in dense urban areas (as in our case studies) have been measured beyond 90% [44].

#### 3.2.2. Mapping Morphologic Slums versus Formal Settlements

We use the *Global Urban Footprint* mapping product that bases on TerraSAR-X and TanDEM-X data [43] providing the classification of 'settlement' areas in our experimental cities. Figure 2c illustrates the respective binary classification. However, this classification does not account for intra-urban structural differences. With respect to the urban poor, the appearances of building structures can be used to delimit their living environments within cities when remote sensing data are applied. We base our classification of 'morphologic slums' on the conceptual ontology suggested by [20] and the empirical study done by [15]: They conceptualize spatial features such as '*highest building density*', '*non-regular, complex alignment of buildings*', '*homogeneity of the pattern*', '*small building sizes*', and '*low building heights*' for delineating areas terminologically named '*morphologic slums*'. Although algorithms automatically deriving these areas in two-dimensional map representations have been developed to a degree where accuracies of 80% and better become possible, e.g., [23,45], we opted for a manual,

visual classification. We do so, as the authors of [10] remark, the visual interpretation still offers the best capability and accuracy for deriving these complex structures.

We follow a standardized digitization protocol to derive a conceptually and spatially consistent morphological slum classification within and across our experimental cities. Based on the VHR optical satellite data, we digitize at a consistent scale of 1:1000 and represent each morphologic slum by a single polygon. The polygons contain several vertices representing real shapes of these neighborhoods. The spatial intersection of the classified ‘morphologic slums’ with the settlement classification derived from the Global Urban Footprint, results in two thematic classes sub-dividing the built environment of the city: ‘morphologic slums’ and ‘formal settlements’. This dichromatic spatial approach serves as spatial proxy for the economic divide in a city.

### 3.3. Mapping the Digital Divide

#### 3.3.1. Twitter Data

Twitter is a web-based message service with 140 characters per ‘tweet’ possible at the time of monitoring (today 280 characters are possible). As of December 2017, *Twitter* has 330 million monthly active users [46]. The tweet contains not solely *semantic contents* by a text message but also additional metadata such as *location* and *time* information.

Our raw data consist of a 1% random sample of the entire population of tweets available through the public API (Application Programming Interface). We spatially defined the API queries using a  $50 \times 50$  km bounding box around each city center. The dataset was streamed over a period of almost one year from 20 June 2016 to 4 June 2017. Ref. [39] find in their study, comparing the data from the streaming API and the Firehose data set (the complete set of data available commercially), that the 1% sample almost returns the complete set of geotagged tweets despite sampling. In consequence we can be highly confident that our geotagged data sets are representative for Twitter users. However, as users can define the (spatial) information they want to share, the number of geotagged tweets is generally comparatively small. In our case, 6.4% of our 1% random sample of tweets carries detailed location information. Other biases integral to the data set are discussed in Section 5.

#### 3.3.2. Filtering and Processing of Twitter Data

Our raw data sample of Twitter messages requires certain preprocessing steps before being suitable for the analysis: For the analysis, we rely only on tweets with an explicit latitude/longitude “point” coordinate. These are 6.4% of tweets from our entire sample.

*First*, we exclude tweets which are spatially associated outside of our two land cover classes (formal settlements and morphologic slums) of interest, i.e., in non-built up areas of our test sites. Urban parks or other places not classified as settlements are excluded (cf. Figure 2a upper left corner).

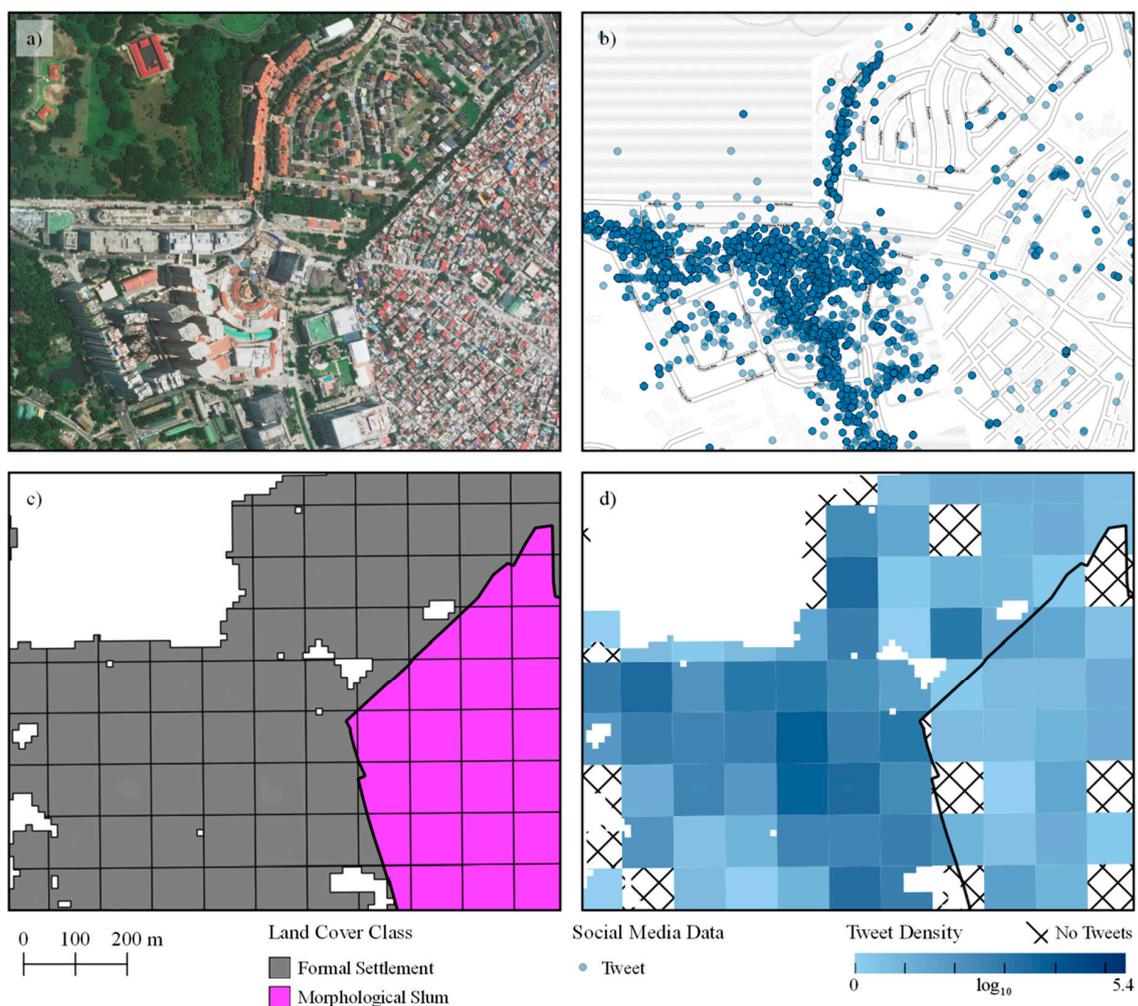
*Second*, a systematic bias occurs when exact geolocations are recorded multiple times. Tweets can be located either by the device’s location (highest quality for this study), check-ins (requires account to be linked with *foursquare* or *path.com*) or by geocoding hashtags or addresses. In the latter case, *Twitter* automatically localizes the tweet. Unfortunately, the recorded twitter data do not include information about how the coordinate pair was generated. Mapping the tweets, however, reveals conspicuous hot spots. As one example, a large number of tweets using the hashtag #*Mumbai*, is located right in front of *Red Cross Street 15*. While this place is central, it is not necessarily a crowded place. Beyond, in contrast to crowded places, where a large number of tweets scatter closely, the coordinates of our example coincide spatially perfectly. Consequently, we assume they were geocoded automatically. In turn, we excluded tweets from the most frequent coordinate pair for each city.

*Third*, a first visualization of the spatial distribution of geolocated tweets reveals some extensive hotspots, possibly induced by *twitterbots* and/or extraordinary users (mainly advertisers). We exclude these hotspots by limiting the maximum number of tweets per user and day to one tweet per spatial

grid unit. Figure 2b illustrates the remaining geolocated tweets of our one-year sample and the emerging spatial pattern.

To capture the variance of twitter activity across space, we use site-specific standard grid geometries. We use a spatial entity of a systematic  $100 \times 100$  m grid. To account for the economic divide within the city, we intersect this systematic grid with the mapped spatial outlines of the two land cover classes ‘formal settlement’ and ‘morphologic slum’. In consequence, the maximum area of a spatial grid unit is 1 hectare when one grid contains only one land cover class. However, the spatial units applied may be smaller when different land cover classes create spatial subunits within the grid. Figure 2c,d illustrate the consistent grid pattern and exemplify the different spatial grid units with respect to varying land cover classes.

Based on the spatial grid units and its subunits we derive twitter activity with respect to location. Therefore, we aggregate the filtered twitter data onto the respective grid unit. We calculate the tweet density per grid unit as the ratio between the total numbers of tweets per hectare. Figure 2d illustrates the aggregated tweet densities.



**Figure 2.** Illustrations of input data processing steps for a subset of Manila: (a) very high resolution (VHR) satellite data © ESRI; (b) Geolocated tweets for an 11-month time period and base map from STAMEN/OSM; (c) Classification of the economic divide by urban morphology characteristics into ‘morphologic slum’ and ‘formal settlement’ and (d) Tweet density aggregated onto a regular grid of  $100 \times 100$ m.

### 3.3.3. Statistical Spatial Variations of Tweet Densities between Morphological Slums and Formal Settlements

In our experiment we aim at analyzing whether systematic differences in tweet density exist between morphological slums and formal settlements. For a consistent analysis across the sampled cities, city specific tweet behaviors need to be eliminated. To do so, we plot the number of tweets per spatial grid unit against the respective spatial extents and fit linear regression models per city through the data. With it, the overall tweet density per individual city can be predicted. This allows eliminating the covariance. For subsequent analyses we rely on the residuals, as they correspond to the tweet density, independent of a city's individual twitter activity characteristics (such as general literacy, access to electricity or local attitude towards the social media platform Twitter). In other words, the residual tweet density represents the ratio of tweets per hectare as if they were tweeted from one generic city.

After the covariance is removed from the data, we conduct an analysis of variance (ANOVA) [47] on the residuals. The ANOVA analyzes the differences among group means and their associated procedures (such as "variation" among and between groups). In consequence, the approach tests if differences are more frequently than random, and thus if they are statistically significant. The variances among averages within a group of data are therefore compared to averages between groups of data. The data groups in our case are the urban land cover classes 'morphologic slums' and 'formal settlements'. The amplitude of the variance between groups measures the differences, with the determination of significance set to 0.01 (1%). For the analysis whether morphological slums or formal settlements feature statistically different tweet densities, we apply an honest significance test (Tukey-Honest Significance Difference Test (Tukey-HSD)). The approach compares the values of the individual data groups to each other. Non-significant differences will be classified as one data group. The confidence level here is set to 99%.

### 3.3.4. Detection of Digital Hot and Cold Spots

So far, our analysis strategy is related to the general differences of twitter activity in morphological slums vs. formal settlements at city level. However, the analysis does not yet account for intra-urban spatial differences. However, as illustrated in Figure 2b, the twitter activity strongly varies on an intra-urban scale. While for the larger proportion of spatial grid units only few if any tweets were observed, others contain up to multiple thousand individual tweets per hectare. Consequently, we aim to detect digital hot and cold spots, i.e., neighborhoods which feature twitter activity significantly above or below the average, and, of course, the digital median of a city.

Using the entire population of spatial grid units per city, we perform a categorical classification of tweet densities. To do so, we use a classification scheme of three quantiles. The quantile embracing the highest tweet density is considered *digital hot spot*. The middle quantile is titled *digital median*, and the lower quantile a *digital cold spot*. Additionally, spatial grid units without any recorded tweets are excluded from this empiricism and straightforwardly constitute a fourth class called *no tweets*. The latter class of *no tweets* is subsequently combined with the lowest quantile to the class of a *digital desert*.

Beyond, we apply spatial statistics based on these tweet density classes, by cumulating the area per land cover class for each land cover class. As the total area and fraction of morphological slums highly varies across the cities (see Table 1), the spatial statistics are normalized by the entire area per land cover class and per city. The resulting relative areas per categorical tweet density class show how frequent each class can be found in morphologic slums as well as in formal settlements. Herewith the spatial share of digital deserts is revealed. The normalized spatial statistics are illustrated in bar charts. We analyze, whether a class, e.g., a 'digital hot spots', is more likely to occur in one of the thematic land cover classes. For significance testing we use a Wilcoxon's rank sum as the data for the tweet density classes are not normal distributed. We test on a 95% confidence level, whether the spatial statistics differs significantly between morphological slums and formal settlements.

### 3.4. Temporal Analysis

All recorded tweets contain the local time when they were sent. This allows in a final step the analysis of temporal patterns comparing morphological slums with formal settlements. Methodically we follow the approach introduced by [34], counting the number of tweets per hour per land cover class, grouped into weekdays and weekends. In contrast to [34] though, we do not use temporal histograms to assess land use types; in contrast, we focus on similarities and differences in temporal behaviors within the land cover classes under investigation. To do so, we aggregate the tweets into temporal signatures for morphological slums as well as formal settlements on a city-wide scale.

In a subsequent step, the temporal trajectories are analyzed in quantitative manner. Therefore, we adapt methods developed in a different discipline—analyzing the phonologic seasonality of time series data, cf. [48]. We assume the daily temporal signatures follow a pattern which can be categorized by a ‘start and an end of a season’. To do so, we *first* fit a cubic spline model through the observed number of tweets per land cover class and hour. We do this for weekdays and weekends separately. Then, *secondly*, the most active phase of a day is extracted based on local extremes of the first derivate. Consequently, the *start of day* (cf. *start of season*) represents the point in time with the largest increment of tweets per hour and vice versa, the largest decrease marks the *end of day* (cf. *end of season*). Assuming that users continuously tweet while using the micro blogging platform, these measures indicate the points in time, when most users come online or go offline respectively. The amount of time between both junctures is considered the *length of day* (cf. *length of season*). Last but not least, we test the hypothesis, that the economic divide also influences the temporal tweet behavior, i.e., the frequency trajectories and its derived metrics.

## 4. Results

In the title of this article we provokingly ask if the “*urban poor are digitally left behind*”? In the following, we approach this question from different perspectives based on our experimental set-up: We present the data and mapping results, we illuminate city-specifics in twitter activities, compare behaviors in morphologic slums vs. formal settlements, and, we analyze whether different temporal signatures between both locations exist.

### 4.1. Data and Mapping Results

Our approach is based on *two main data products*: (1) The classifications of ‘morphologic slums’ and ‘formal settlements’ for mapping the economic divide within the complex urban landscape, and (2) the twitter data streamed over a period of almost one year.

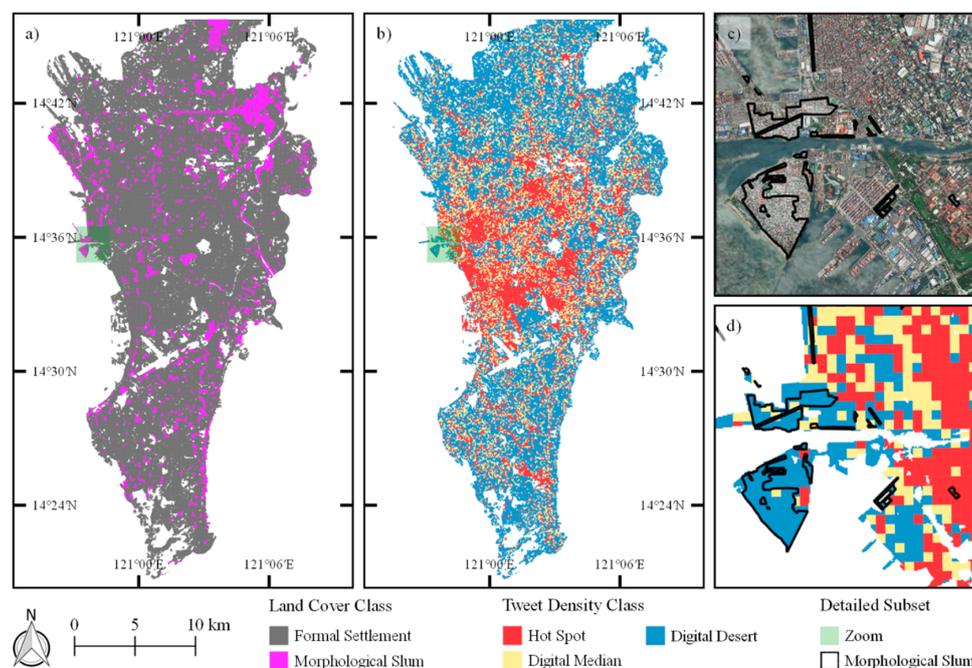
1. In general, we classify 274,184 hectare of cumulated settlement areas for all our sample cities. We find that only 5.54% of the settlement areas are occupied by ‘morphologic slums’. However, for the area share of morphologic slums of the total settlement area we observe a strong variation across cities from 18.90% in Caracas to only 0.11% in Lisbon (Table 1).
2. The preprocessed twitter data set contains 3.73 million geolocated tweets cumulated for all sample cities. We find that twitter activity varies significantly across cities, with Manila featuring more than 2 million tweets vs. Dhaka with only about 30,000 tweets within the time period of monitoring. Beyond, we also find that a relatively small share of tweets of 2.7% is localized in morphologic slum areas. However, we also observe strong variations of shares of tweets in morphologic slums across cities from 8.74% in Caracas to only 0.07% in Lisbon (Table 1).

**Table 1.** Detailed spatial classification results of settlement areas, spatial shares of morphologic slums, number of recorded tweets and tweets in morphologic slums and related numbers of users.

City	Settlement Area Based on Remote Sensing Data		Number of Recorded Tweets and Shares of Tweets		Number of Twitter Users	
	Total (ha)	Morphologic Slum (%)	Total (n)	Morphologic Slum (%)	Total (n)	Morphologic Slum (%)
Dhaka	20,817	6.79%	28,395	3.11%	17,773	3.1%
Mumbai	23,940	14.21%	133,161	3.5%	79,540	2.99%
Manila	47,552	6.09%	2,038,719	1.61%	1,265,235	1.36%
Caracas	16,333	18.9%	233,876	8.74%	108,346	7.81%
Rio	36,191	6.66%	999,525	4.55%	583,973	4.15%
Cairo	52,482	2.53%	122,192	0.43%	71,325	0.3%
Capetown	40,673	1.49%	87,779	0.64%	50,962	0.47%
Lisbon	36,196	0.11%	150,861	0.07%	98,136	0.04%

In a first general observation, we find that relative to the spatial shares comparatively fewer tweets are sent from the locations of morphological slums (Table 1). This discrepancy becomes even stronger when looking at the number of unique usernames compared to both—area and number of tweets. Consequently, the latter measurements indicate that the ratio of tweets per user is higher in morphological slums. Thus, we assume the twitter activities in formal settlements are based on a broader user base.

We exemplify the land cover mapping results and the tweet density distribution for the city of Manila. Figure 3a illustrates the administrative extent of the city, the morphologic slum and formal settlement classification. From a general point of view, we observe morphologic slums located dispersedly across the entire city. Figure 3b displays the tweet density at grid level, as it was classified in Section 3.3.3. We observe a decrease in twitter activity in Manila from central areas along the western waterfront to peripheral areas north and south.



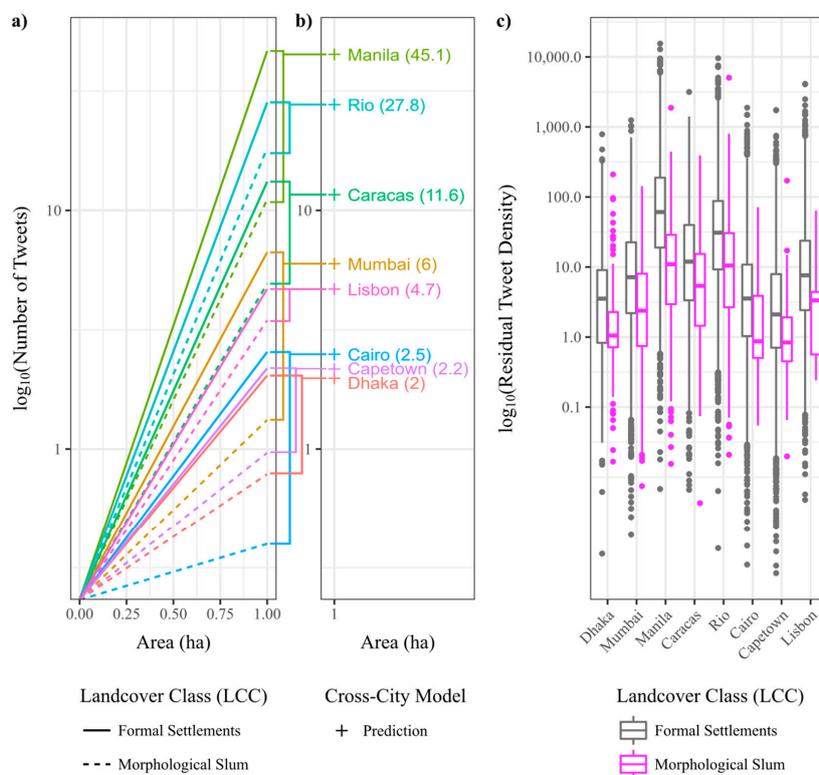
**Figure 3.** Mapping results for Manila: (a) Classification of the urban landscape into two categorical classes ‘formal settlement’ and ‘morphological slum’; (b) Tweet density class projected onto the grid; Detailed subset showing (c) settlement structures at the western waterfront and (d) corresponding tweet density class.

#### 4.2. Variance Analysis of Tweet Densities Across Cities and Within Cities (between Morphological Slums and Formal Settlements)

The cities under investigation feature very different social media activities (cf. Table 1). To account for the city-characteristic local specifications, we use linear models to eliminate this heterogeneity across the various sampled cities as a covariance a priori. With it, we find city specific usage quantities for the two land cover classes morphologic slums and formal settlements (Figure 4a) as well as for the entire cities in general (Figure 4b).

We find that the number of tweets per spatial unit is on average consistently lower in morphologic slum areas than in formal settlements across all cities. In Manila, as one example, the number of tweets per spatial grid unit is 3.72 times smaller as in formal settlements (Figure 4a). However, taking a cross-city perspective, we find for example the tweet density in morphologic slums in Manila significantly higher than in formal settlements of Cairo. In consequence, we find every city has unique city-specific usage quantities of the twitter social media platform. Analyzing the tweet densities across all study cities reveals Manila and Rio de Janeiro are the most active online cities of our sample. The cross-city model predicts 45.1 (Manila) and 27.8 (Rio de Janeiro) geo-referenced tweets per hectare. In contrast, the twitter activity in the other cities is significantly lower. With 2.5 or even less tweets per hectare, the cities of Capetown, Cairo and Dhaka have the least overall tweet density (Figure 4b).

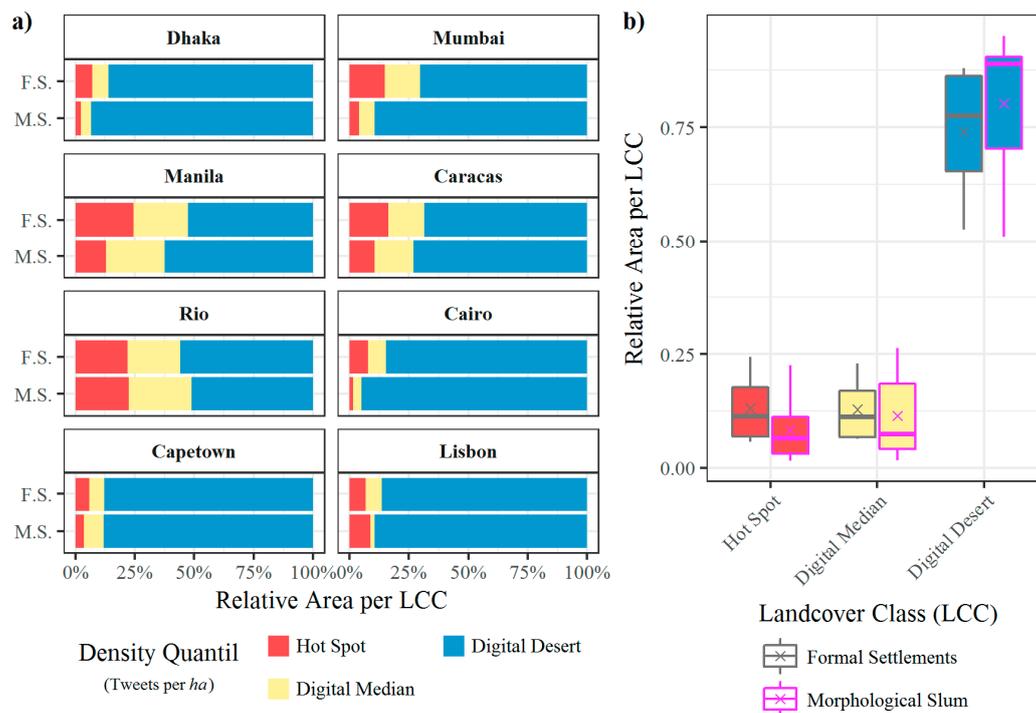
In the following, we proceed with the residual tweet density to reduce effects of the revealed city-specific twitter activities of the sample cities. In general, we find that in all eight cities, the median residual features lower tweet densities within morphological slums compared to formal settlements. Beyond, the boxplot's central quantiles are shifted towards a lower tweet density (Figure 4c). Consequently, the Tukey's HSD following the ANOVA identified a highly significant difference in twitter activity between morphological slums and formal settlements ( $p$ -value < 0.0001).



**Figure 4.** Empirical statistics on tweet densities, as well as the cross-city model residuals: (a) Linear models illustrating tweet density depending on city and land cover class; (b) Cross-city model combining both land cover classes used to extract overall tweet densities for each city; (c) Residual tweet density after city-specific covariation was removed.

### 4.3. Spatial Statistics of Digital Hot Spots and Cold Spots

We highlight spatial concentrations of twitter activity by using the classification of digital hot spots, digital medians and the semantic digital desert class. In our analysis we relate the classification scheme to our thematic urban land cover classes, morphologic slums and formal settlements, by summarizing spatial statistics. In the following bar charts, the urban areas covered by the particular digital activity class are cumulated (Figure 5a). The length of each bar is scaled relative to the total area. Figure 5b summarizes these zonal statistics for each digital activity class over all cities.



**Figure 5.** Zonal statistics of tweet density classes per city and land cover class: (a) Bar charts showing the fraction of areas covered by the tweet density classes. As size of cities and especially morphological slums vary (see Table 1), the areas are normalized to 100%, thus each bar cumulates to a total area covered by the respective land cover class. The abbreviation F.S. represents formal settlements and M.S. represents morphological slums; (b) Boxplots summarizing the zonal statistics by digital activity class. The mean is additionally displayed with a cross.

In general, we find that only small shares of areas in every city can be considered digitally oriented (Figure 5). Even for the most digitally oriented cities of Manila and Rio de Janeiro, we measure 53.2% and 55.5% of the total areas being classified as digital deserts (*cold spots* and *no tweets* combined). For all other cities, it is found that digital deserts cover at least 69.4% of the areas as in the case of Caracas. In Mumbai (72.9%), Cairo (84.9%), Lisbon (86.5%), Dhaka (86.7) and Cape Town (87.9%) the shares of digital deserts are even higher.

In detail, it is also worth noting that digital deserts are predominantly found in morphological slums. As example, in Mumbai we find digital deserts in *morphological slums* (89.5%) 1.3 times more likely than in *formal settlements* (70.2%). This trend is confirmed for the cities of Manila (1.2 times as likely), Cairo (1.1), Dhaka (1.1) and Caracas (1.1). For Lisbon (1.0) and Cape Town (1.0) the relation is basically even and for Rio de Janeiro (0.9), however, this trend cannot be confirmed as morphological slums feature a relatively lower share of digital deserts. Overall, these numbers thus hint at a general trend that urban poor are digitally left behind.

We find statistical differences between morphological slums and formal settlements by comparing medians. In formal settlements on average 11.4% of the area are classified as digital hot spot, opposed to 6.5% in morphological slums. For the other tweet density classes, this discrepancy is weaker—11.2% vs. 7.4% concerning digital medians and is inverse for areas covered by digital deserts (88.8% in morphological slums versus 77.4% in formal settlements).

4.4. Temporal Signatures of Twitter Activities

In general, we find residents in morphologic slum areas are less digitally oriented than in formal settlements. In the following we investigate whether among the twitter users’ differences in temporal behaviors exist. Therefore, we integrate the number of tweets per land cover class on an hourly basis.

For the temporal analysis, however, the quantity of tweets per time unit (one hour) reduces the basic population of available twitter data. Thus, the least digitally oriented cities of Lisbon, Cairo, Dhaka and Cape Town are disregarded for the temporal analysis, as for morphological slums not enough data are existent within the one-year time period of monitoring.

For the four remaining cities with an adequate share of tweets per hour, we identify two phases in temporal signatures: The first phase involves the morning hours before noon. Here the tweet frequency increases sharply and reflects the start of online activity. The second phase involves the afternoon hours with the daily maximum of twitter activity around 6 p.m. Both phases are predominantly separated by a downturn around noon. It is important to stress though, that the amplitude of the downturn varies. As example, this is especially evident in Caracas; however, the “lunchbreak” is also measured at a high activity level, leading to a plateau shape in Rio de Janeiro and Manila. When comparing weekdays to weekends, we observe in general similar temporal signatures; however, a temporal delay of twitter activity on weekends is revealed (Figure 6).

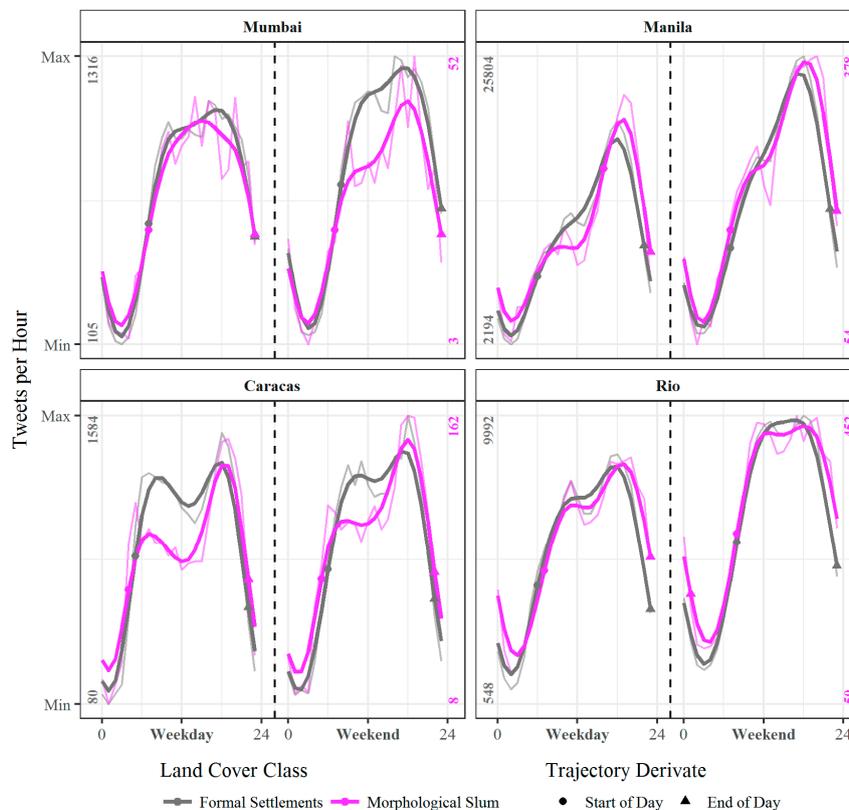


Figure 6. Temporal signatures for weekdays and weekends for each city and land cover class under investigation. Bold, smooth lines are spline fitted models over lighter colored raw data (ragged lines). Points and triangles mark trajectory derivate.

When we relate the measured temporal signatures of twitter activity to the economic divide—morphological slums vs. formal settlements—we find that the general temporal signatures are similar. In detail, we reveal that relative tweet frequencies in morphological slums are slightly higher at night and in the late evening hours compared to formal settlements. Over the course of the day, however, the number of tweets in morphologic slums is measured lower, which is particularly significant during the first phase of the day.

From an intra-urban perspective—when comparing extracted benchmarks of the two land cover classes defining the economic divide—we find no significant difference among user groups (Table 2). Only slight differences with a temporarily longer average *length of day* in formal settlements (15.85 h) compared to morphological slums (15.36 h) are measured ( $p\text{-Value}_{\text{LOD}} = 0.708$ ). Beyond, the metrics of the start and end of day do not differ significantly ( $p\text{-Value}_{\text{SOD}} = 0.477$ ;  $p\text{-Value}_{\text{EOD}} = 0.212$ ) between the user groups defining the economic divide. In consequence, we find that among the twitter users in morphological slums their temporal behavior is similar to the twitter users in formal settlements.

**Table 2.** Trajectory derivatives and root-mean-squared error (RMSE) of the spline fitted temporal signature. SOD = Start of day; EOD = End of day; LOD = Length of day.

		Weekday				Weekend			
		SOD	EOD	LOD	RMSE	SOD	EOD	LOD	RMSE
Mumbai	Formal Settlement.	7.08	23.06	15.98	0.039	8.09	23.06	14.97	0.042
	Morphological slum	7.08	23.06	15.98	0.073	7.08	23.06	15.98	0.080
Manila	Formal Settlement.	6.07	22.04	15.98	0.032	7.08	22.04	14.97	0.033
	Morphological slum	16.04	23.06	7.01	0.044	7.08	23.06	15.98	0.056
Caracas	Formal Settlement.	5.06	22.04	16.99	0.052	6.07	22.04	15.98	0.048
	Morphological slum	4.04	22.04	18.00	0.058	5.06	22.04	16.99	0.057
Rio	Formal Settlement.	6.07	23.06	16.99	0.035	8.09	23.06	14.97	0.028
	Morphological slum	7.08	23.06	15.98	0.038	8.09	1.08	16.99	0.043

## 5. Discussion

The World Migration Report [5] stresses, that although we are living in times where more data are available than ever before, we are facing a massive lack of data on urban poverty. This study explores, whether the combination of two different data sets—from *remote sensing* and *social networks*—allows for reducing the knowledge gaps on urban poverty.

Picking up the provocative question in the title of this study, we find that the economic divide influences digital participation in public life. Fewer residents of morphological slums are found to produce data on the twitter platform, so overall this social group appears to be less digitally oriented (“*digitally left behind*”), compared to residents of formal settlements. At the same time, we find that among the twitter user the temporal behavior is similar in morphologic slums and formal settlements. These empirical findings only become possible with a comparatively large sample of cities at city-wide scale, as the combination of these data sets allows extending the capabilities of individual disciplines for research towards urban poverty. These main results need to be discussed from the perspective of the limitations of our input data sets: The *first limitation* applies to working with remotely sensed data. Our mapping approaches for morphological slums and formal settlements result in very precise, high resolution data sets delineating the economic divide within cities. However, the economic divide assessed by built-up structures is only a proxy for describing different social groups. As [15] reveal, urban poor are also located in other structural types, such as in high rise facilities and vice versa. In consequence, our proxy for mapping poor populations only comprises a subset of the targeted social group. However, as socio-economic data are scarce, the remotely sensed approach is, as it has been argued by [25] and [49], a legitimate and consistent one across cities. Additionally, a typical issue concerning remote sensing is the difference between land use (related to activities on the ground) and

land cover (related to surface structure). In this study, that difference introduces a bias concerning the formal settlement class. The global urban footprint includes residential, as well as industrial areas. When comparing the digital divide between morphological slums and formal settlements the following issue is existent: the poor areas are mainly characterized by residential land use whereas formal settlements are characterized by mixed land uses. This issue is illustrated in the detailed subset of Figure 5.

The second limitation refers to twitter data: We utilized these data for describing the digital divide. However, it is clear, that twitter is neither representing all online activity nor all internet users, as not everybody with internet access also participates in social networks (or specifically our subset of the Twitter platform). Further, studies of [30] as well as [41] have shown that especially Twitter is populated by young, male, highly educated users. Being aware of these circumstances, we consider our twitter data a proxy only. In contrast to actual internet service provider (ISP), the utilized network data may even come with some advantages. Qualitative interviews with social network users in an African slum revealed, that most users living in poor conditions are hesitant to close contracts bound to fixed monthly fees: “[ . . . ] Like most users in developing countries, [they] use pre-paid airtime or credit to access the mobile Internet” [32] (p. 2825). Further, acquiring and homogenizing all available ISP data from the eight sample cities, was outside the scope of our study. Twitter, however, being used on all inhabited continents of our planet [29], embodies one consistent data source. Hereby the freely available 1% random sample is found to be representative for the firehose, especially when using geotagged tweets [39]. However, the 1% random sample of Twitter data also poses methodological challenges. Although overall 113 million Tweets were streamed, after the preprocessing, the numbers of georeferenced tweets were reduced to a basic data population of 3.7 million. This is still a large amount of data; however, in some less digitally oriented cities such as Lisbon, the basic population, especially for the small spatial shares of morphological slums, resulted in a relatively small amount of twitter data. By putting these numbers in spatial relation, this allows identifying cities where the social network of Twitter is significantly less popular, but it generates statistical challenges due to a too low basic population of data for our streaming period. This is especially true for the analysis of temporal signatures, where the limited data volume relies on a statistical model. In addition, we are aware that the smartphone’s location accuracy of geotagged tweets is around several meters [50]. Last but not least, we want to amend that our research design ignores the individual mobility of people over the course of a day. As for example, inhabitants of morphological slums may work in formal settlements. Thus, they influence tweet quantities related to other land cover classes. In consequence, the combination of both data sets may inherit an unknown error of spatial misclassification to the respective thematic classes representing the economic divide. Last but not least, the heterogeneity of our city sample did not reveal any relationship between the size of a city and its twitter characteristics.

In spite of these underlying assumptions and challenges related to the input data sets, we find this approach legitimate in the absence of better and more up-to-date data sets. Although these proxies come with individual limitations, they allow comparing multiple sites in a globally consistent way. And, combining both data sets allows reducing knowledge gaps on urban poor: We find that—as we ask provocatively in the title—that the urban poor are digitally left behind to a certain degree. We reveal that digital hot spots are found fewer in morphological slums and digital deserts dominate here. However, the current study still relies only on eight cities and a one-year period of twitter data. It therefore is important to apply this methodology with a larger sample of data.

In this study we relate the twitter activity relative to the area. [41], however, suggest that the identification of digital cold spots is more appropriate when relating social media activity to population density. This is especially relevant, as the mapped land cover classes feature very different population densities. Unfortunately, population density at the spatial level of morphological slums is not available or not even existing for most cities in the Global South. Other studies have, however, consistently shown, that population densities in morphological slums are mostly significantly higher than in other

parts of the city e.g., [7,51,52]. In consequence, it is very likely that the trend of morphological slums being less digitally oriented is even stronger than measured here.

Last but not least, we transferred time series analyses on tweet frequency trajectories. We did so, as our data feature similar characteristics as phonological data with respect to its seasonality. Although this approach reached its capabilities for weekdays in the morphological slums of Manila, where the first increasing tweet activity phase of the day is significantly less intense than the second one. Thus, we measure the start of day here for 16:04. With this exception, the derived metrics allow comparing the tweet frequency charts and temporal patterns in an objective manner. Consequently, we propose to add these methods to the domain of social analyses.

## 6. Conclusions

This study combines two different data sets—*remote sensing* and *twitter*—and reveals that the economic divide within cities reflects different digital participation in public life. The analysis supports the assumption that the city is not only spatially divided by the appearances of morphologic building characteristics but also by an invisible digital divide (“*the poor are digitally left behind*”).

*First*, it is found that participation in modern communication techniques such as social networks is comparatively scarce for people living under precarious conditions compared to the city average. The number of tweets in morphological slums is generated by a smaller number of users. However, it needs to be noted that exceptions to this general trend have been identified. *Second*, most of the morphologic slum areas are classified as digital deserts; and it is found vice versa, digital hot spots are predominantly classified in formal settlements. *Third*, it is revealed that among the twitter users very similar temporal behavior patterns over the course of a day exist at both sides of the economic divide.

We conclude that studies are in demand that allow confirming or declining these results with better data than the proxies used in this study. Beyond, extending the analyses empirically by the number of sample cities and the twitter data volume is in demand. And, last but not least, we suggest that analyzing context information of the tweets will provide further insights into social group’s behavior and thinking.

**Author Contributions:** Conceptualization, Methodology and Writing-Original Draft, H.T., J.S.; Formal Analysis, J.S.; Writing-Review and Editing, M.W., C.G.; Supervision, S.D.; Funding Acquisition, H.T., X.Z.

**Funding:** This research was funded by the European Research Council (ERC) under the European Union’s Horizon 2020 research and innovation program, grant number 714087-So2Sat. We thank European Space Imaging (EUSI) for providing high resolution optical satellite data.

**Conflicts of Interest:** The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, and in the decision to publish the results.

## References

1. Hollis, L. *Cities Are Good for You: The Genius of the Metropolis*, 1st ed.; Bloomsbury Press: New York, NY, USA, 2013; ISBN 978-1-62040-206-1.
2. UN-Habitat. *Planning Sustainable Cities: Global Report on Human Settlements 2009*; Routledge: London, UK, 2009; ISBN 1-134-90071-6.
3. UN-Habitat. *Urbanization and Development Emerging Futures: World Cities Report*; United Nations Publication: New York, NY, USA, 2016.
4. Milanovic, B. *Die Ungleiche Welt: Migration, das Eine Prozent und Die Zukunft der Mittelschicht*; Suhrkamp: Berlin, Germany, 2016.
5. International Organization for Migration. *World Migration Report 2015: Migrants and Cities: New Partnerships to Manage Mobility*; International Organization for Migration (IOM): Geneva, Switzerland, 2015; ISBN 978-92-9068-709-2.
6. Tacoli, C.; McGranahan, G.; Satterthwaite, D. *Urbanisation, Rural-Urban Migration and Urban Poverty*; International Institute for Environment and Development: London, UK, 2015; ISBN 978-1-78431-137-7.

7. Taubenböck, H.; Wurm, M. Ich weiß, dass ich nichts weiß—Bevölkerungsschätzung in der Megacity Mumbai. In *Globale Urbanisierung*; Springer: Berlin, Germany, 2015; pp. 171–178.
8. Crampton, J.W.; Graham, M.; Poorthuis, A.; Shelton, T.; Stephens, M.; Wilson, M.W.; Zook, M. Beyond the geotag: Situating 'big data' and leveraging the potential of the geoweb. *Cartogr. Geogr. Inf. Sci.* **2013**, *40*, 130–139. [[CrossRef](#)]
9. OECD. Poverty Rate (Indicator). 2018. Available online: <https://doi.org/10.1787/0fe1315d-en> (accessed on 24 July 2018).
10. Sliuzas, R.; Mboup, G.; de Sherbinin, A. *Report of the Expert Group Meeting on Slum Identification and Mapping*; CIESIN, UN-Habitat, ITC: Enschede, The Netherlands, 2008; Volume 36.
11. Bibi, S. Measuring Poverty in a Multidimensional Perspective: A Review of Literature. PEP Working Paper No. 2005-07. Available online: <https://ssrn.com/abstract=850487> (accessed on 1 November 2005).
12. Harvey, D. *Rebellische Städte: Vom Recht auf Stadt zur Urbanen Revolution*; Suhrkamp: Berlin, Germany, 2013; ISBN 978-3-518-12657-8.
13. Barber, B.R. *If Mayors Ruled the World: Dysfunctional Nations, Rising Cities*; Yale University Press: New Haven, CT, USA, 2014; ISBN 978-0-300-20932-7.
14. Glaeser, E.L.; Kahn, M.E.; Rappaport, J. Why do the poor live in cities? The role of public transportation. *J. Urban. Econ.* **2008**, *63*, 1–24. [[CrossRef](#)]
15. Taubenböck, H.; Kraff, N.J.; Wurm, M. The morphology of the Arrival City—A global categorization based on literature surveys and remotely sensed data. *Appl. Geogr.* **2018**, *92*, 150–167. [[CrossRef](#)]
16. Perlman, J.E. Marginality from Myth to Reality—The Favelas of Rio de Janeiro 1968–2005. 2005. Available online: [http://advantronsample2.com/Marginality\\_from\\_Myth\\_to\\_Reality.pdf](http://advantronsample2.com/Marginality_from_Myth_to_Reality.pdf) (accessed on 5 June 2018).
17. Mitra, S.; Mulligan, J.; Schilling, J.; Harper, J.; Vivekananda, J.; Krause, L. Developing risk or resilience? Effects of slum upgrading on the social contract and social cohesion in Kibera, Nairobi. *Environ. Urban.* **2017**, *29*, 103–122. [[CrossRef](#)]
18. Wratten, E. Conceptualizing urban poverty. *Environ. Urban.* **1995**, *7*, 11–38. [[CrossRef](#)]
19. Kuffer, M.; Pfeffer, K.; Sliuzas, R. Slums from Space—15 Years of Slum Mapping Using Remote Sensing. *Remote Sens.* **2016**, *8*, 455. [[CrossRef](#)]
20. Kohli, D.; Sliuzas, R.; Kerle, N.; Stein, A. An ontology of slums for image-based classification. *Comput. Environ. Urban Syst.* **2012**, *36*, 154–163. [[CrossRef](#)]
21. Owen, K.K.; Wong, D.W. An approach to differentiate informal settlements using spectral, texture, geomorphology and road accessibility metrics. *Appl. Geogr.* **2013**, *38*, 107–118. [[CrossRef](#)]
22. Engstrom, R.; Sandborn, A.; Qin, Y.; Burgdorfer, J.; Stow, D.; Weeks, J.; Graesser, J. Mapping slums using spatial features in Accra, Ghana. In *Proceedings of the Urban Remote Sensing Event, 2015 Joint, Lausanne, Switzerland, 30 March–1 April 2015*; pp. 1–4.
23. Wurm, M.; Taubenböck, H.; Weigand, M.; Schmitt, A. Slum mapping in polarimetric SAR data using spatial features. *Remote Sens. Environ.* **2017**, *194*, 190–204. [[CrossRef](#)]
24. Sandborn, A.; Engstrom, R.N. Determining the Relationship between Census Data and Spatial Features Derived From High-Resolution Imagery in Accra, Ghana. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2016**, *9*, 1970–1977. [[CrossRef](#)]
25. Wurm, M.; Taubenböck, H. Detecting social groups from space—Assessment of remote sensing-based mapped morphological slums using income data. *Remote Sens. Lett.* **2018**, *9*, 41–50. [[CrossRef](#)]
26. Norris, P. *Digital Divide: Civic Engagement, Information Poverty, and the Internet Worldwide*; Cambridge University Press: Cambridge, UK, 2001.
27. Hersberger, J.A. Are the economically poor information poor? Does the digital divide affect the homeless and access to information? *Can. J. Inf. Libr. Sci.* **2003**, *27*, 44–63.
28. Blank, G. The Digital Divide among Twitter Users and Its Implications for Social Research. *Soc. Sci. Comput. Rev.* **2017**, *35*, 679–697. [[CrossRef](#)]
29. Leetaru, K.; Wang, S.; Cao, G.; Padmanabhan, A.; Shook, E. Mapping the global Twitter heartbeat: The geography of Twitter. *First Monday* **2013**, *18*, 5–6. [[CrossRef](#)]
30. Li, L.; Goodchild, M.F.; Xu, B. Spatial, temporal, and socioeconomic patterns in the use of Twitter and Flickr. *Cartogr. Geogr. Inf. Sci.* **2013**, *40*, 61–77. [[CrossRef](#)]

31. Nemer, D. Online Favela: The Use of Social Media by the Marginalized in Brazil. *Inf. Technol. Dev.* **2015**, *22*, 364–379. [[CrossRef](#)]
32. Wyche, S.P.; Forte, A.; Yardi Schoenebeck, S. Hustling online: Understanding consolidated facebook use in an informal settlement in Nairobi. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Paris, France, 27 April–2 May 2013; pp. 2823–2832.
33. Klotz, M.; Wurm, M.; Zhu, X.X.; Taubenböck, H. Digital deserts on the ground and from space—An experimental spatial analysis combining social network and earth observation data in megacity Mumbai. In Proceedings of the Joint Urban Remote Sensing Event, Dubai, UAE, 6–8 March 2017; pp. 1–4.
34. Frias-Martinez, V.; Soto, V.; Hohwald, H.; Frias-Martinez, E. Characterizing Urban Landscapes Using Geolocated Tweets. In Proceedings of the 2012 International Conference on Privacy, Security, Risk and Trust and 2012 International Conference on Social Computing, Amsterdam, The Netherlands, 3–5 September 2012; pp. 239–248.
35. Domínguez, D.R.; Díaz Redondo, R.P.; Vilas, A.F.; Khalifa, M.B. Sensing the city with Instagram: Clustering geolocated data for outlier detection. *Expert Syst. Appl.* **2017**, *78*, 319–333. [[CrossRef](#)]
36. Liu, Y.; Liu, X.; Gao, S.; Gong, L.; Kang, C.; Zhi, Y.; Chi, G.; Shi, L. Social Sensing: A New Approach to Understanding Our Socioeconomic Environments. *Ann. Assoc. Am. Geogr.* **2015**, *105*, 512–530. [[CrossRef](#)]
37. Cranshaw, J.; Schwartz, R.; Hong, J.I.; Sadeh, N. The Livelihoods Project: Utilizing Social Media to Understand the Dynamics of a City. In Proceedings of the Sixth International AAAI Conference on Weblogs and Social Media, Dublin, Ireland, 4–7 June 2012; pp. 58–65.
38. Goodchild, M.F. The quality of big (geo)data. *Dialogues Hum. Geogr.* **2013**, *3*, 280–284. [[CrossRef](#)]
39. Morstatter, F.; Pfeffer, J.; Liu, H.; Carley, K.M. Is the Sample Good Enough? Comparing Data from Twitter’s Streaming API with Twitter’s Firehose. In Proceedings of the Seventh International Conference on Weblogs and Social Media, Cambridge, MA, USA, 8–11 July 2013.
40. Mislove, A.; Lehmann, S.; Ahn, Y.-Y.; Onnela, J.-P.; Rosenquist, J.N. Understanding the Demographics of Twitter Users. In Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media, Barcelona, Spain, 17–21 July 2011; p. 11.
41. Anselin, L.; Williams, S. Digital neighborhoods. *J. Urban. Int. Res. Placemak. Urban. Sustain.* **2016**, *9*, 305–328. [[CrossRef](#)]
42. GADM. *Database of Global Administrative Areas*; GADM: Montreal, QC, Canada, 2018.
43. Esch, T.; Taubenböck, H.; Roth, A.; Heldens, W.; Felbier, A.; Thiel, M.; Schmidt, M.; Müller, A.; Dech, S. TanDEM-X mission—New perspectives for the inventory and monitoring of global settlement patterns. *J. Appl. Remote Sens.* **2012**, *6*, 061702. [[CrossRef](#)]
44. Klotz, M.; Kemper, T.; Geiß, C.; Esch, T.; Taubenböck, H. How good is the map? A multi-scale cross-comparison framework for global settlement layers: Evidence from Central Europe. *Remote Sens. Environ.* **2016**, *178*, 191–212. [[CrossRef](#)]
45. Graesser, J.; Cheriadat, A.; Vatsavai, R.R.; Chandola, V.; Long, J.; Bright, E. Image Based Characterization of Formal and Informal Neighborhoods in an Urban Landscape. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2012**, *5*, 1164–1176. [[CrossRef](#)]
46. Twitter Inc. *Annual Report*; Twitter Inc.: San Francisco, CA, USA, 2017.
47. Chambers, J.; Freeny, A.; Heiberger, R. Analysis of variance; designed experiments. *Stat. Model. S* **1992**, *5*, 145–193.
48. Filippa, G.; Cremonese, E.; Migliavacca, M.; Galvagno, M.; Forkel, M.; Wingate, L.; Tomelleri, E.; Morra di Cella, U.; Richardson, A.D. Phenopix: A R package for image-based vegetation phenology. *Agric. For. Meteorol.* **2016**, *220*, 141–150. [[CrossRef](#)]
49. Engstrom, R.; Hersh, J.; Newhouse, D. *Poverty from Space: Using High-Resolution Satellite Imagery for Estimating Economic Well-Being*; Policy Research Working Papers; World Bank Group: Washington, DC, USA, 2017.
50. Zandbergen, P.A. Accuracy of iPhone Locations: A Comparison of Assisted GPS, WiFi and Cellular Positioning. *Trans. GIS* **2009**, *13*, 5–25. [[CrossRef](#)]
51. Davis, M. *Planet. der Slums*; Assoziation A: Berlin, Germany, 2007; ISBN 978-3-935936-56-9.
52. Fuchs, M. Slum als Projekt: Dharavi und die Falle der Marginalisierung. In *Mumbai–Delhi–Kolkata. Annäherungen an die Megastädte Indiens*; Draupadi: Heidelberg, Germany, 2006; pp. 47–63.

