Applied Geography 181 (2025) 103687

Contents lists available at ScienceDirect

Applied Geography

journal homepage: www.elsevier.com/locate/apgeog

The digital urban frontier: Disparities in social media activity between consolidated and newly urbanized areas in Africa

Johannes Mast [a,*], Marta Sapena [a], Henri Debray [b], Justice Nana Inkoom [c], Richard Lemoine-Rodríguez [b], Christian Geiß [d], Hannes Taubenböck [a,b]

[a] German Aerospace Center (DLR), Münchener Straße 20, 82234, Weßling, Germany
[b] Julius Maximilian University of Würzburg, Am Hubland, 97074, Würzburg, Germany
[c] Martin Luther University Halle-Wittenberg, Von-Seckendorff-Platz 4, 06120, Halle (Saale), Germany
[d] Rhenish Friedrich Wilhelm University of Bonn, Meckenheimer Allee 166, 53115, Bonn, Germany

1. Introduction

In the 21st century city, urban space is both physical and digital. Terms such as "smart city", "digital twin", and "digital citizenship" illustrate that urban communities are increasingly complemented by information-rich overlays and digitally represented (Robertson & Feick, 2016) on social networks, web-mapping sites and content-sharing platforms. The spread of information & communication technology and geospatial software applications make the digital space increasingly important for how individuals and organizations perceive the city and, by extension, make decisions within it (Törnberg & Uitermark, 2022). In this way, the power to shape the cultural, social, and economic reality of urban areas is partly dependent on the visibility they have in the digital sphere. But this visibility is not evenly distributed, neither among citizens nor among the spaces in which they live (Boy & Uitermark, 2017; Zhu & Lerman, 2016). The competition for attention has led to the development of entire industries and technologies around an "urban attention economy" (Törnberg, 2023). Online visibility is a form of power, and it is in limited supply: not every place can be the center of attention. But which places are?

The spatial distribution of online visibility on social media can be found in explicit form on web platforms that offer a geolocation functionality, such as Twitter (now X) or Instagram. Compared to other social media services like Facebook or WhatsApp, these platforms are notable because they allow users to explicitly link their contribution to specific geolocations, creating geolocated social media (GSM) data. The result is a layer of GSM activity that covers cities like a perceptible "digital skin" (Rabari & Storper, 2015) – a skin that exists in both virtual and geographic space and has the potential to connect the two worlds. As Kelley (2011) describes it, GSM are positioned to fill the gaps between the material landscape and the socio-cultural facts of how we perceive, experience, and interact with this environment. For this reason, they are frequently used sources of information for individuals and indicators for places' positions in the urban attention economy. Importantly, online visibility can be approximated via GSM activity (the number of GSM content pieces attached to a certain location) and reception (number of responses, typically in the form of likes, shares, or replies).

The distribution of GSM activity is highly heterogeneous across and within cities (Robertson & Feick, 2016). Understanding variation across urban space can support urban studies in several ways: *Firstly*, it helps identifying types of intraurban areas for which GSM data is promising for research and planning. *Secondly*, understanding what drives differences in GSM activity can reveal digital disparities and underlying socio-cultural inequalities (Lemoine-Rodríguez, Mast, et al., 2024). Previous works identified inequalities in social media use along demographic lines such as gender, ethnicity, and age (Hargittai, 2020; Malik et al., 2015; Wentrup et al., 2016). In geographic space, disparities exist between countries (Huang & Carley, 2019), within countries (Sanderson et al., 2024), and even within cities, between more and less prosperous parts of neighborhoods (Taubenböck et al., 2018) and from centers to peripheries (Jiang et al., 2016). However, one aspect of the digital urban divide was not researched yet, despite its increasing relevance considering the rapid growth rates of many cities: The digital disparity between older and newer parts of settlements.

New settlements are no negligible part of the urban landscape: As humanity is undertaking its largest migration ever – from rural to urban areas (Saunders, 2010) – cities are growing at unprecedented rates due to a combination of natural population increase, migration, and the transition of formerly rural to urban areas (UN DESA 2019). The way in which cities accommodate the newly arriving population is mostly through the creation of new built-up areas (henceforth NBA), usually close to existing older built-up areas (henceforth OBA). Although population densities are also increasing in OBAs, NBAs absorb the majority

* Corresponding author. German Aerospace Center (DLR), Münchener Straße 20, 82234, Weßling, Germany.
E-mail address: Johannes.mast@dlr.de (J. Mast).

https://doi.org/10.1016/j.apgeog.2025.103687
Received 26 December 2024; Received in revised form 2 May 2025; Accepted 2 June 2025
Available online 11 June 2025
0143-6228/

of urban population growth. Analyses of satellite imagery over the past decades have shown that the urban extent of many cities has increased at astonishing speed and in a wide variety of patterns (Taubenböck et al. 2020, 2024, 2025). This diversity is the result of differentiated paths of historical processes that directly impacted urban developments, such as demographic growth, colonization (Şen, 2024), economic growth, and changes in planning culture (Taylor, 2013). Even within the extent of a same city, differences in urban morphology testifies to these historical processes. Trying to summarize them in a general and reproduceable way, Debray et al. (2023) showed that the connection between historical processes and intra-urban morphology can be captured through the concept of intensity of plannedness (henceforth IoP). The concept of IoP summarizes the combined historical processes influencing intra-urban layout on a gradient of how forcefully it has been planned. This gradient is conceptualized in five categories ranging from very spontaneous urban developments resulting from the decisions of many individuals to extremely planned development resulting from the manicured planning of a single institution or estate company.

While the underlying processes are manifold, they affect settlements' form and density, which are related to socio-economic factors and the quality of life of their inhabitants (Debray et al., 2023; Sapena et al., 2020).

For all these reasons, NBAs across the world, even when sharing a same development date, can be very heterogeneous in terms of morphologies. Therefore, NBAs cannot be associated with a specific and unique morphological type of urban fabric: they stand on their own as a spatiotemporal frame in the dynamic process of urbanization. NBAs are characterized by two factors: 1) They are new i.e., they consist of recently constructed buildings; 2) They are places where many (or all) new residents have only recently settled. While we cannot assume anything about the distance people moved to relocate to these newly built housing areas, this still makes them places of comparatively high population dynamism, and thus an interesting unit of analysis for migration research. And it is the first characteristic, their recency, that allows their large-scale analysis through remote sensing time series. The temporal granularity of satellite data supports fine-grained analyses of settlement age which have revealed the massive extent to which cities have grown in recent decades (Lemoine-Rodríguez et al., 2020; Liu et al., 2020; Taubenböck et al., 2024). While the distribution of GSM in the urban space has been analyzed in several studies (e.g., Taubenböck et al., 2018; Jiang et al., 2016; Cai et al., 2017; Lv et al., 2021; Robertson & Feick, 2016; Yin and Guangqing, 2021; Lang et al., 2022; for reviews of the use of Twitter data in urban geography, see Zhu et al., 2022; Smith et al., 2025), to our knowledge the link between GSM activity and settlement age remains unexplored.

This is especially relevant on the African continent, which has, together with Asia, the highest urban growth rates (Kamana et al., 2024; Taubenböck et al., 2024) and between 2020 and 2050 is predicted to double its population from 704 million to 1.4 billion people by 2050 and increase its total urban footprint from 175,000 km^2 to 450,000 km^2 (OECD, 2025). Because of this, Africa can be considered the place where we find the world's urban frontier: A region where new territory is transformed into urban space at unparalleled pace and at the same time, much urban potential yet to be realized (Goodfellow, 2022). Compared to elsewhere, much of Africa's urban expansion is unregulated (Güneralp et al., 2017), with substantial heterogeneity in urban morphology (Taubenböck et al., 2020). At the same time, a high degree of urban primacy indicates that it is traditional urban centers that still dominate most African countries, with disproportionately less resources and attention given to other urban areas (Güneralp et al., 2017), suggesting a divide that may also extend into the digital sphere. In other words, the *digital* urban frontier may lag behind the physical urban frontier. Whether this disparity exists, and whether regional characteristics or the plannedness and structure of settlements influence it, has not been investigated so far.

In this study, we address this research gap. Our hypothesis is that the settlement age affects how GSM is used, influenced by a possibly increased affinity of new residents to discuss their new environment, NBAs' novelty providing initial advantages in terms of economic attention, or, potentially, disadvantages resulting from delayed development of cultural, social or economic places and telecommunication infrastructure. To validate our hypotheses, we address the following research questions:

How does the density of GSM activity differ between new and old areas of African settlements? If imbalances exist, do they differ in frequency and magnitude (2) between regions of Africa, (3) between spontaneous and planned settlements, and (4) based on the spatial structure of NBAs?

By identifying if –and where– a digital divide can be measured in Africa's cities, we aim to supply the knowledge base on urban development with a digital perspective to enable effective urban strategies (Kamana et al., 2024) that adequately incorporate all parts of urban agglomerations in a digital transformation (African Union Commission, 2021; OECD, 2025).

2. Data acquisition and processing

2.1. Study period and study sites

We conducted this study for a ten-year period between January 2010 and December 2019. We selected this timespan due to the availability of a consistent set of geolocated social media posts from the platform Twitter where the geotagging feature became first widely used in 2009 and substantially changed in 2019, reducing availability of fine-grained geotags (Kruspe et al., 2021, pp. 212–221). To identify the age of all built-up areas that already existed before this study period or were built-up during the study period, we used remote sensing data to map the built-up year not just for the study period, but as far back in time as data allows (see Section 2.3). It was not possible to analyze the entire African continent due to data limitations. While remote sensing data is available, access to social media data was less abundant. Consequently, we conducted a site-selection process to identify 135 study sites of 5 km radius (Fig. 1) where the most urban area was added between 1985 and 2019, according to the World Settlement Footprint dataset (Marconcini et al., 2020). Focused on capturing any areas where strong growth occurred, the sites are not intended to reflect any particular urban boundaries and their centroids often do not correspond to a city center. Additionally, we ensured the representation of different geographic and cultural contexts (Fig. 1) by stratifying sites across United Nations (UN) georegions (United Nations Statistical Office, 1982), selecting at least 20 sites for each georegion. A description of the site-selection process and the site outlines can be accessed in Supplement A.

2.2. Built-up expansion

For each site, we used remote sensing data to map the yearly expansion of built-up area. Due to the absence of frequent long-term data on built-up expansion in Africa, we extended the World Settlement Footprint Evolution (WSF_{Evo}) dataset by Marconcini et al. (2021) which provides settlement extents from 1985 to 2015, using the Do-it-yourself built-up (DIY-BU) mapping tool (Sapena et al., 2024) for the years 2015–2019. Unlike WSF_{Evo}, which is derived from Landsat imagery at 30 m resolution and global scale, DIY-BU uses Sentinel imagery to produce local urban expansion maps at 10 m resolution. Compared to global multi-temporal products, its locally fine-tuned approach using building footprint data has been shown to provide more accurate results (for more details, see Sapena et al., 2024). We resampled the WSF_{Evo} to match the spatial resolution of the DIY-BU maps. Through spatial overlap, we assigned the earliest date from WSF_{Evo} to the built-up pixels in DIY-BU maps, resulting in urban expansion maps covering 1985 to 2019 at 10 m resolution (Fig. 2).

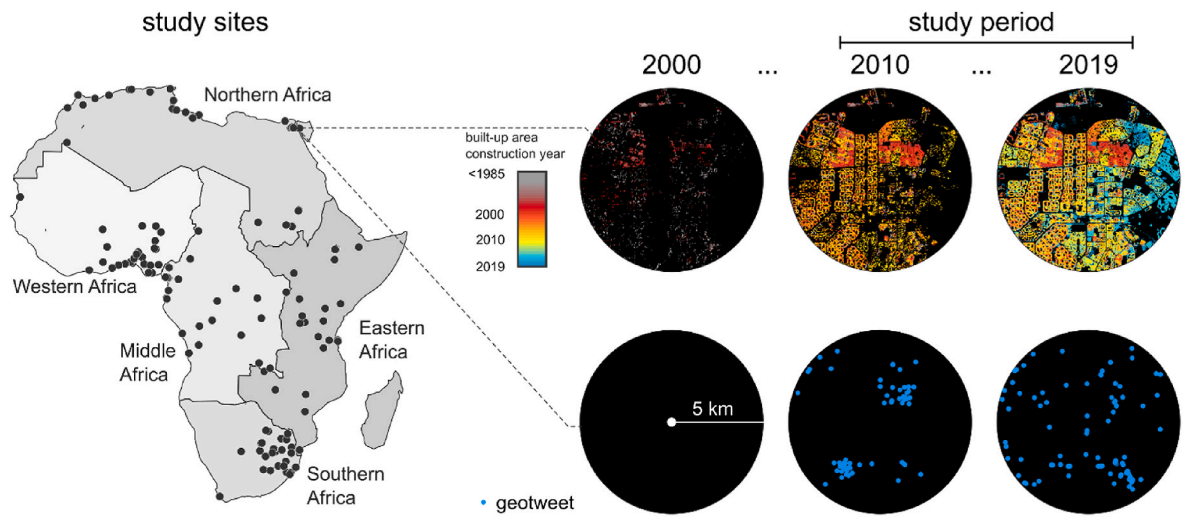


Fig. 1. Overview of the 135 study sites (left) and the primary data sets: geotweets (lower right) and built-up areas with construction year (upper right), mapped by the World Settlement Footprint Evolution (WSF_{Evo}) and the Do-it-yourself built-up (DIY-BU) mapping tool.

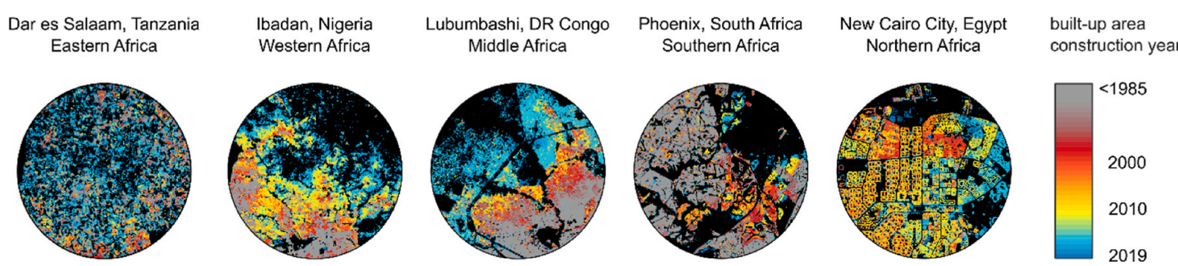


Fig. 2. Examples of urban expansion in study sites.

2.3. Twitter data

As a source for GSM data, we used geolocated tweets (geotweets) from the microblog platform Twitter, a favored data source in many fields because of its coverage and widespread use (Karami et al., 2020). Twitter users may attach geolocation in several ways, primarily through place-based tags of various precision levels (point-of-interest, neighborhood, city, administrative, and country), or through precise coordinates obtained from the device or third-party applications (Kruspe et al., 2021, pp. 212–221). We were granted access to Twitter's API via Twitter's Academic Research Track (Twitter Inc, 2022), which allowed for the query of 10 million tweets per month via V2 of the Twitter API. We used the R package academictwitteR (Barrie & Ho, 2021) in December 2022 and January 2023 to download all geotweets available for the study period (1. January 2010–1. January 2020) and bounding boxes of the study sites. From this set, we kept those that we could precisely locate within the circular boundary of a site via a point-of-interest-level place tag or precise coordinates. If a geotweet contained both, we geolocated the tweet using precise coordinates.

The availability of sufficiently precise geotweets varied over time, likely due to changes in Twitter's API, applications, and policies, or user preferences. Of 14,391,264 geotweets initially returned by the query, we discarded those with insufficient spatial precision (coarser than point-of-interest-level) and excluded accounts producing many geotweets with clear evidence of automation (unusually high user activity, speed, and concurrency of tweets, compare Petutschnig et al., 2020; Lemoine-Rodríguez, Mast, et al., 2024; Mast et al., 2024). This resulted in a final total of 6,997,346 geotweets for the 135 sites. No metadata was recorded other than account ID and the calendar year in which each geotweet was posted. We integrated the geotweet and remote sensing datasets by annotating each geotweet with the construction year of the built-up pixel at its location. Geotweets posted from non-built-up pixels were excluded from the analysis.

3. Analysis strategy

After the datasets were thusly integrated, the disparity in geotweet activity between OBAs and NBAs could be addressed analytically. The simple distinction between OBA and NBA permitted limited nuance, thus, we instead treated built-up age as a continuum. Along this continuum, the disparity in geotweet density was quantified and evaluated with significance tests at the site-level to answer research question 1 (Fig. 3). Subsequently, research questions 2–4 were tested by relating the significance and magnitude of the disparity to geographic regions, site-level planning, and spatial metrics of settlement structure, as explained in sections 3.4–3.6.

3.1. Geotweet density

We sought to attain a general understanding of the relationship between built-up age and geotweet activity in our study sites. Therefore, we combined information on geotweets and on the spatial dynamic of settlement expansion. We operationalized this for the geotweets by weighting them by their impacts. We structured each AOI into a spatial grid of 100-m-wide square cells and computed the weight for each geotweet as:

$$W_{tweet} = \frac{1}{n_{per\ user}} + n_{retweets} + n_{replies}$$

with $n_{per\ user}$ being the amount of geotweets posted by a given account within the same month in the same grid cell, $n_{retweets}$ being the number of times the geotweet was re-tweeted and $n_{replies}$ the amount of replies the

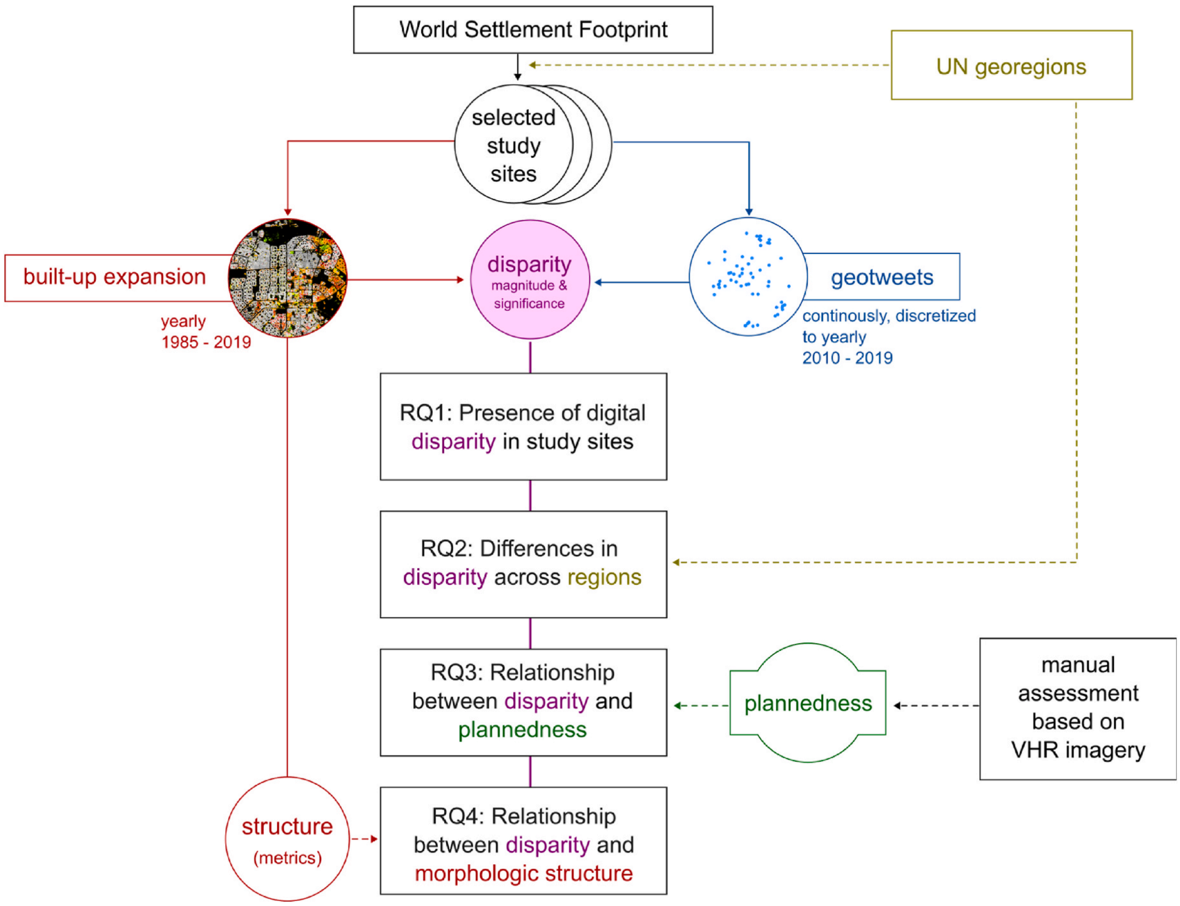


Fig. 3. Schematic workflow of the study.

tweet generated. In this way, we reduced the local impact of highly active users and shifted the focus from the geotweets' production to their visibility, assuming that a retweet, quote, or reply of a geotweet increases the digital representation of an area just like a new geotweet would.

From the geotweet data, we associated the weight of the geotweet, its geolocation and its publication date. With this, from the built-up expansion data, we retrieved the date of construction of the built-up environment of the tweets' geolocation to compute the age of the built-up environment at the publication date of each tweet. After aggregating this information by publication year of the geotweets and by age of the built-up environment, we computed the weighted density of geotweets as:

$$density_X^{age} = \frac{Ntw*_X^{age}}{area_X^{age}}$$

with $Ntw*_X^{age}$ being the count of tweets adjusted by their weights for the year X (between 1985 and 2019) for the category of built-up age of their geolocation, denoted *age*. Accordingly, $area_X^{age}$ is the surface occupied by the built-up environment of the category of built-up age *age* as of the year X of the geotweet publication. In other words, for each site, we computed the distribution of the weighted geotweet counts across built-up age of the settlement in the period 1985–2019 (see Fig. 4).

3.2. Quantifying disparity

This procedure yielded a sequence of geotweet density values for each site and year, which can be seen as a distribution of density by age of construction (Fig. 4f). Our first research question concerned the balance of this distribution, which we hypothesized to be skewed towards the new or old areas. In order to compare this age-related digital disparity across study sites, we designed a new measure. This measure is similar to skewness metrics and represents the degree to which the measured density disparity approaches the theoretically maximum possible disparity. We refer henceforth to this measure as the digital density skew (DDS) and describe its computation in detail in Supplement B where we also compare it with alternative measures. The DDS is bounded between −1 and 1, respectively corresponding to the cases where all tweets posted for a given year are located only in the newest (DDS = −1) or the oldest parts (DDS = 1) of a settlement (in the year the tweets were posted). Accordingly, a value of 0.5 indicates that the disparity is half of its theoretical maximum. The value of zero indicates balanced geotweet density between older built-up areas and newer built-up areas of the site (relatively to the other values in the observed sequence). While this interpretation of new and old built-up as relative to the observed sequence is not necessarily appropriate in the historical context of every site, it is consistent and comparable. Fig. 5 illustrates how the DDS responds to various distributions.

After calculating the DDS separately for each of the 10 calendar years of the study period, we calculated the overall DDS for each study site as the mean DDS over all years, excluding years that did not contain at least 100 geotweets produced by at least 10 different accounts. Further, we excluded study sites that did not contain tweets by at least 100 different accounts overall.

Additionally, we tested the robustness of the DDS towards GPS-inaccuracies, low numbers of geotweets, low numbers of users, and the impact of highly active accounts (Supplement C).

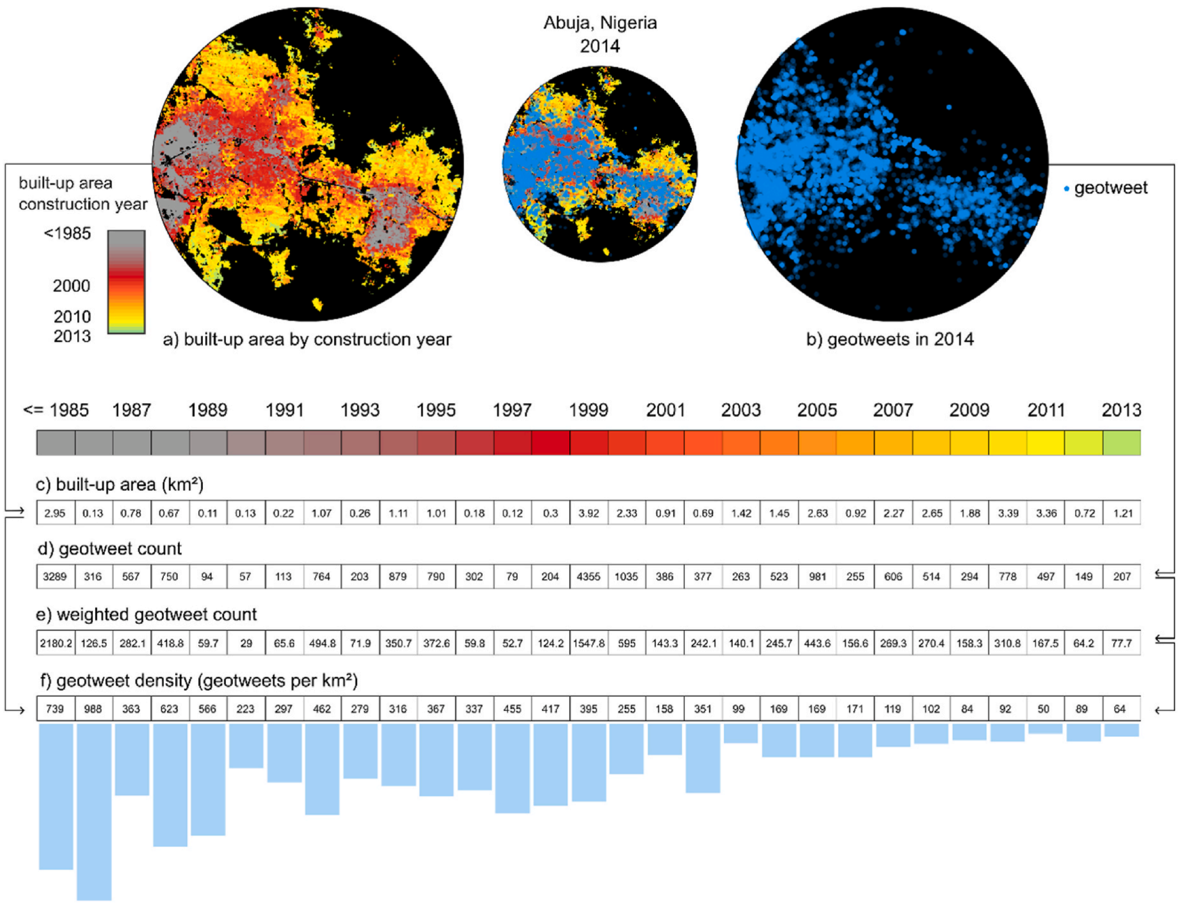


Fig. 4. Derivation of density distribution for the year 2014 in a single site (Abuja, Nigeria). Items (a) and (b) display the two input data sets used to compute the built-up area (c) and the geotweet count (d & e) per construction year, which are joined to compute the geotweet density across construction years, where the geotweet density is higher in older areas (left) than newer areas (right).

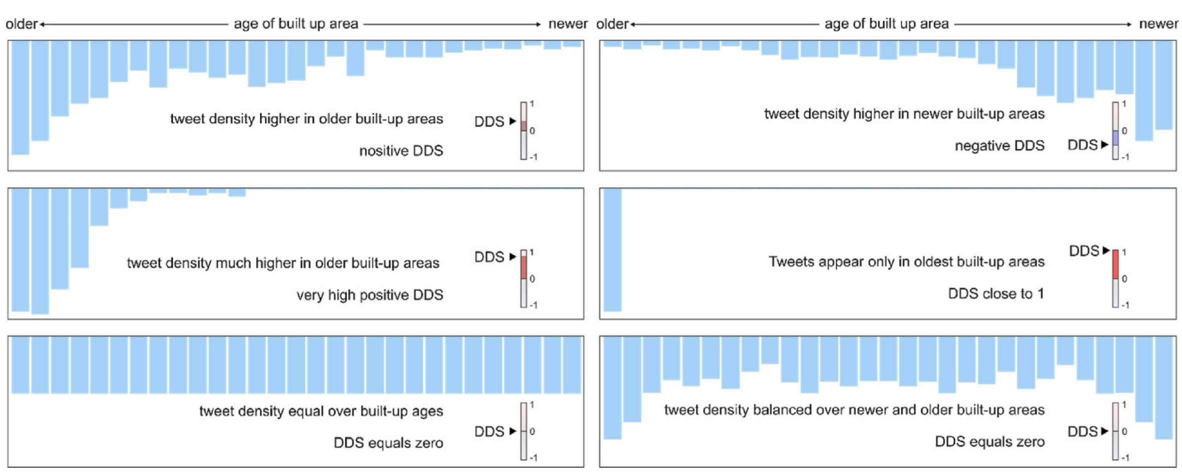


Fig. 5. Several fictive tweet density distributions (compare Fig. 4 above) and their quantification by the disparity measure digital density skew (DDS).

3.3. *Testing disparity at site level*

The magnitude of the DDS indicates the degree to which there is an observed association between age and geotweet activity within a site, but it does not indicate its statistical significance.

To assess the significance, we compared the DDS against a null distribution. Because the weighting approach and topologic effects might affect new areas differently than older ones, the probability distribution of the DDS is not necessarily symmetric and centered on zero. Thus, to create a null distribution that accounts for these effects, we used a permutation test (Supplement D). We used the R package spatstat to model the spatial distribution of geotweets for each site as a clustered point process with a Cauchy distribution (Baddeley & Turner, 2005). Unlike a completely random spatial sampling, modeling the geotweet distribution as a clustered point process reflected empirical observations about the distribution of tweets (Steiger et al., 2015).

From the model, we simulated 1000 spatial point patterns for each site, each having a spatial overall clusteredness similar to the original distribution, but with different locations of clusters. Reassigning the geotweets to these new coordinate-pairs, we repeated the process of extracting the built age, the geotweet density, and the DDS, resulting in 1000 reference DDS values which approximate a null-distribution of DDS values for each site. On this basis, we assessed the significance of the DDS via a right-tailed p-value (fraction of permutations whose DDS value was at least as high as the observed DDS) and a left-tailed p-value (fraction of permutations whose mean DDS value was at least as low or lower than the observed mean DDS), applying a significance threshold of 95 %.

A positive DDS indicated that geotweet activity was denser in older parts of settlements and a significant right-tailed test for the DDS indicated that this is higher than could be caused by chance. Conversely, a negative DDS along with a significant left-tailed test indicated that the activity was significantly denser in newer settlements.

3.4. Comparing across African regions

We tested whether digital disparity varies significantly across geographic regions. We followed the UN geoscheme to group sites into five geographic regions: Northern Africa, Eastern Africa, Middle Africa, Western Africa, and Southern Africa (United Nations Statistical Office, 1982; Fig. 1). These regions are one of many possible geographical stratifications, and like any, cannot accurately represent all cultural, historical, or ethnic divisions. Acknowledging this limitation, we still considered these regions the best suiting available systematic division scheme to the best of our knowledge, corresponding to United Nations standards, and being widely used for statistical purposes.

We applied both parametric and non-parametric ANOVAs of the DDS values over the five regions. The non-parametric ANOVA is a robust alternative that does not, like the parametric version, assume normality (assessed with a Shapiro-Wilk test (Royston, 1982; Razali and Yah 2011) and heterogeneity of variances across groups (assessed with a Levene, 1960). We report p-values and ω^2 and η^2 as effect sizes for the parametric and non-parametric ANOVA, respectively. We also applied post-hoc two-sample t-tests to identify between which regions differences can be found.

We further compared prevalence (number of sites) and frequency (proportion of sites) of significant disparity for each region.

3.5. Comparing across intensity of plannedness

According to the intensity of plannedness (IoP) ontology (Debray et al., 2023), we assessed the IoP of each site using experts' knowledge supported by visual inspection of up-to-date aerial imagery, OpenStreetMap data and Google Street View imagery.

The assessment was firstly performed individually by the authors and, in a subsequent step, differences between the authors' assessments were discussed until an agreement was reached. Each site was assigned one of the five IoP degrees ranging from completely spontaneous to completely planned at block level and at street-level separately with respect to the morphologic and structural appearance in the image data. Fig. 6 shows examples for each of the five IoP degrees.

To test RQ3, similarly as for RQ2, we performed both nonparametric and a parametric ANOVA of the DDS by the five IoP degrees.

3.6. Modeling the relationship between disparity and the spatial structure of settlements

Finally, we assess whether digital disparity varied significantly according to the spatial structure of the built-up area in each site. This spatial arrangement can indicate socioeconomic processes that may affect digital disparities but are challenging to analyze directly due to a lack of data. Thus, to explore whether any links exist between spatial structure and digital disparities, we used spatial metrics to quantify the spatial structure of built-up areas in the sites. We calculated the metrics for the built-up area in 2019. The selected metrics have been shown to successfully quantify urban morphology (Sapena et al., 2020). These metrics focus on the size, shape, distribution, and connectivity of built-up areas. Metrics such as mean patch size (MPS), urban density (UD), and object density (OD) describe the extent and concentration of built-up patches, while the shape index (SI) and urban compactness (UC) capture the complexity and cohesion. Measures like the Euclidean nearest-neighbor distance (ENND) and area-weighted standard distance (AWSD) describe the spatial arrangement and centrality of built-up patches. Additionally, the porosity index (PI), effective mesh size (EMS), and dispersion index (DI) quantify fragmentation, openness, and distribution patterns, offering a comprehensive view of how built-up areas are organized within the sites (see Supplement E for a detailed description of the metrics).

Additionally, we calculated the spatial partial information (SPI) component of entropy, as described by Altieri and Cocchi (2022) in the analysis of urban fragmentation patterns. Compared to the spatial metrics suggested by Sapena et al. (2022), the SPI can be calculated at arbitrary scales, parametrized by distance ranges, and it is inherently scaled between 0 and 1, allowing us to compare built-up compactness at various scales. A high SPI, especially at short distance ranges, corresponding to fine scales, hints at compact urban structure. Using the R

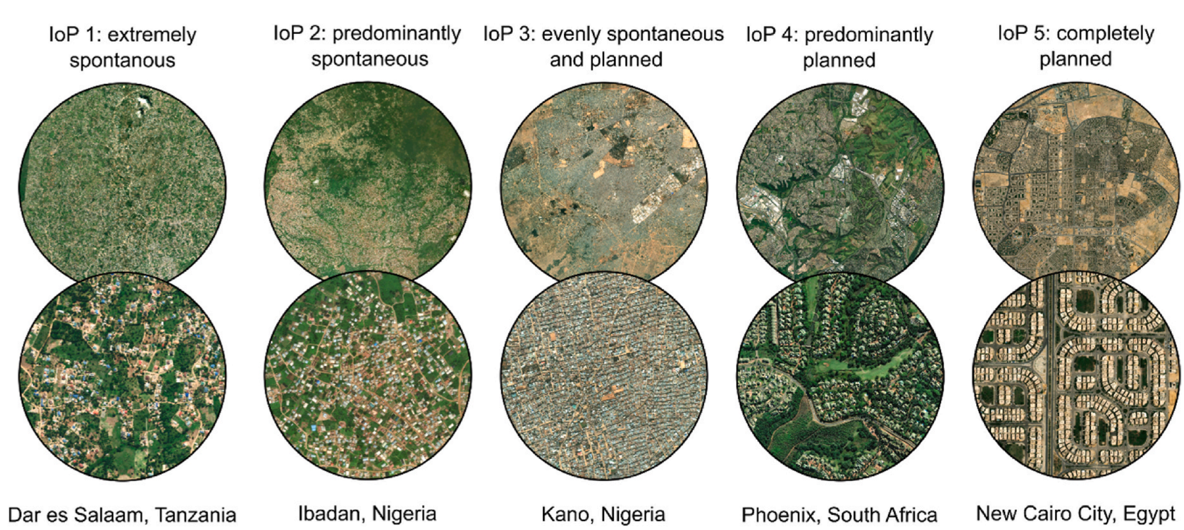


Fig. 6. Example sites with varying intensity of plannedness (IoP) (similar for both street-level and block level). Data: Microsoft (Bing Maps).

package SpatEntropy (Altieri et al., 2021), we calculated the SPI at four scales based on the distance-ranges: less than 100 m; 100–500 m; 500–2000 m; 2000 m and more.

To test for a relationship between the DDS and the spatial metrics, we fitted a series of linear regression models between DDS and each metric. We report the significance of the relationships and the coefficients of determination (R^2) of these models. We fitted three different models. First, we used all sites regardless of the significance of the DDS to model the one-to-one relationship between the DDS and the spatial metrics and assess their relationship. Second, we repeated this only for the sites with significant DDS and explored the intensity of the relationship. And third, as a single metric can only represent a single aspect of urban form, we fitted a multiple regression model to uncorrelated metrics using a forward stepwise regression approach to receive a simpler model with only the significant metrics.

4. Results

4.1. Disparity at site level

We quantified the digital disparity using the DDS measure to test whether geotweet density varies by built-up age (RQ1). Overall, the mean DDS was measured at 0.201 (standard deviation = 0.16), significantly higher than the mean of any of the 1000 permutations (right-tailed p-value = 0.000), making it very unlikely to result from spatial randomness. Similarly, the mean DDS for significant sites was measured with 0.308, i.e., 30.8 % of the theoretical maximum disparity. Both perspectives support the hypothesis of an overall trend towards disparity between OBA and NBA. With the mean DDS being significantly higher than the distribution of mean DDS under the null, the results provide unambiguous evidence of a widespread, but not uniform, disparity (see examples in Fig. 7).

At the level of individual sites, 63 of 135 showed a significant right-tailed test and a positive DDS, indicating higher digital density in OBA. For 55 sites, the result was not significant, and 17 sites were excluded because twitter data was insufficient for most years. Few sites were measured with a negative DDS, and their left-tailed tests were not significant, such as *Zango, Angola* (−0.272), *Lumpundu, Congo* (−0.217), or *Al-Fashir, Sudan* (−0.195), suggesting that there were no cases where geotweet activity was higher in newer areas. Results for all sites are listed in Supplement F.

The sensitivity analyses indicate that the reported results of the significance tests are robust towards realistic magnitudes of GPS inaccuracy and the impact of highly active accounts (see Supplement C for details).

To complement these overall statistics, we provide detailed description of sites which exemplify different cases. Firstly, an example of a site where twitter activity is far denser in OBAs is *Nador, Morocco* (DDS = 0.367, p = 0.01, B2 in Fig. 7). This mediterranean city is surrounded by a lagoon to the east and mountains on the other sides. The majority of Nador's built-up area originated before the start of our study period and forms a consistently settled core area. Recent expansion occurred predominantly in A) the western suburbs in the form of scattered rows of connected houses, or in B) a highly planned area on the south-eastern shore, where many hotels and restaurants suggest a touristic focus. While this new planned area contains several hundred of the city's geotweets, the aforementioned suburbs contain very few. The vast majority of the 17 K tweets were distributed throughout the core, with geotweet density increasing along the main roads and declining towards the edges.

Ginti, Nigeria (B1 in Fig. 7*)*, is the site with the largest settlement expansion in our study. A moderate disparity was measured here (DDS = 0.257, p = 0). Located in Ikorodu, north-east of Lagos, the site displays a continuous settlement expansion that extends outwards from the main road and is only roughly bounded by green spaces and wetlands. In contrast to Nador, almost all expansion occurred in recent decades. The roughly 62 K geotweets were distributed throughout all the built-up area, but most densely in the central area near the roads, which are spaces of social and commercial activity. Notable was a single concentration of 17 K geotweets in one location close to the center of the site, but from different accounts and with different content. While the concentration of tweets along the core areas resembles Nador, many of the core areas in Ginti are more recent, leading to a weaker measurement of disparity.

In *Sheikh Zayed City, Egypt (B3 in* Fig. 7*)*, west of Cairo, large neighborhoods were developed in recent decades in a comparatively brief time and highly planned manner. The 110 K geotweets were distributed fairly evenly across the built-up area, with notable concentrations in malls (the two largest built around 2010) and a high-tech business district in the north (built around 2000). As these sites were built-up throughout the study period, no significant disparity was measured for this site (DDS = −0.093, p = 0.85).

Bani Walid, Libya (B4 in Fig. 7*)* is an oasis town divided by a valley into a northern and a southern half. Despite heavy fighting in the aftermath of the Libyan civil war in 2012 (Gumuchian, 2012), the town has expanded in south-western, south-eastern, and northern direction. Although the built-up area has doubled during the study period, almost all of the 2.5 K geotweets are located in the OBAs. There, they were highly concentrated, and in some cases, dozens of tweets were attached to the same precise coordinate. This phenomenon was occasionally observed in other sites, but particularly impactful here due to the comparative sparsity of tweets. Coupled with the low number of users, this led to a high variance of the permuted disparity. Consequently, the measured disparity, although moderately high, was not significant (DDS = 0.281, p = 0.22).

4.2. Disparity across African regions

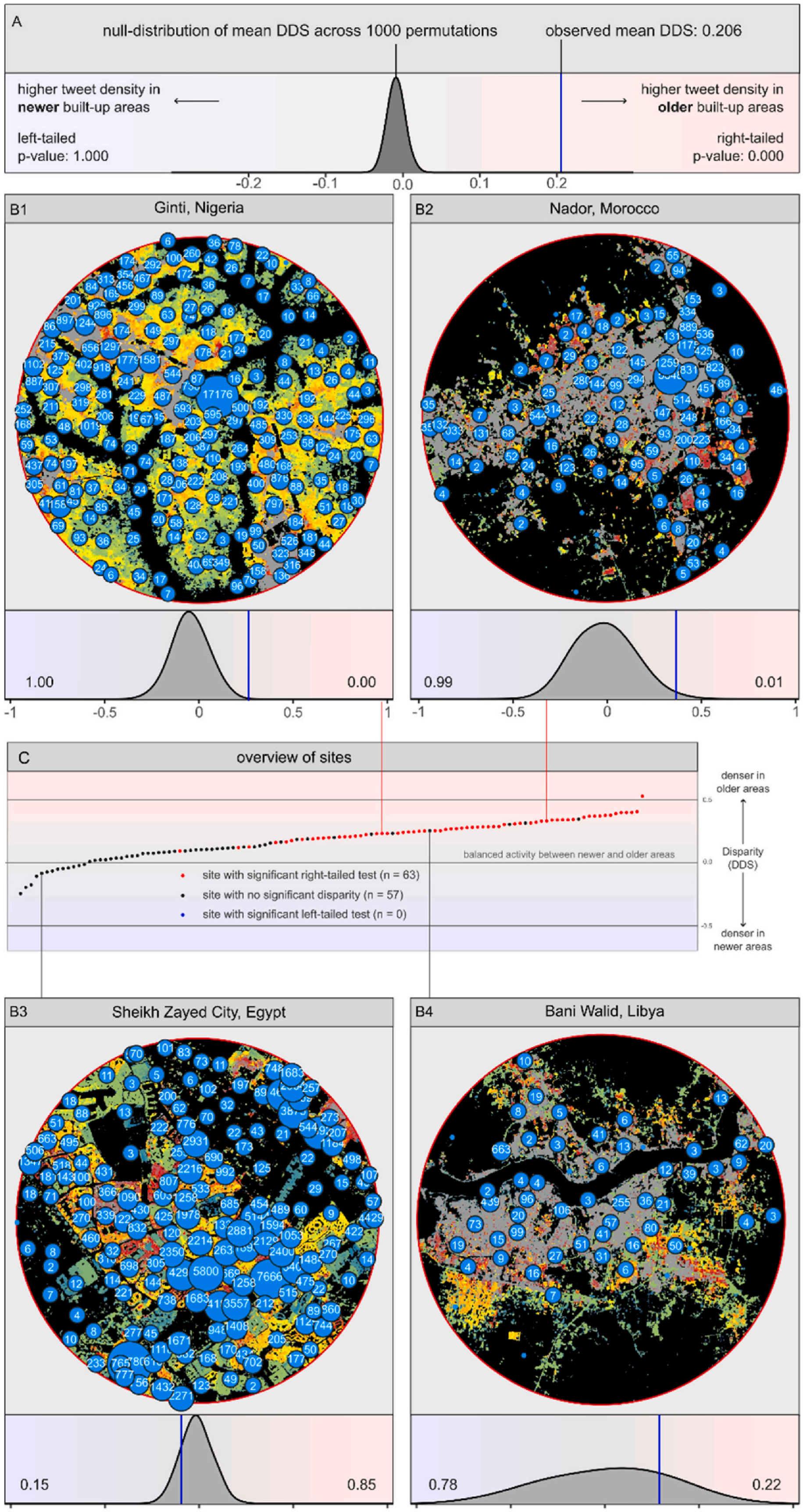
Both parametric and non-parametric ANOVA tests for geographical differences in digital disparity were significant (with p-values of 0.002 and 0.001, respectively), at a moderately high effect size ($\omega^2 = 0.108$ and $\eta^2 = 0.1295$, respectively). There were no extreme outliers, and the distributions of sites' DDS were approximately normal (Shapiro-Wilk test p-value>0.05) within each of the five regions, but evidence of heterogeneity of variances was found across regions (Levene-test p-value = 0.012). Altogether, this provides compelling evidence of differences between some, but not all, regions.

Examining the frequencies illustrated in Fig. 8, we found that the frequency of disparity varies by region. In Western Africa, the DDS was significant for most sites (23 of 29). In Northern and Southern Africa, the majority of sites did not display significant DDS. In addition, data availability varied across regions. In Eastern and Middle Africa, more than a fifth of sites provided insufficient data, meaning that in no year, more than 100 geotweets from at least 10 accounts were available.

4.3. Disparity across degrees of IoP

Regarding the digital disparity between different levels of spontaneous and planned settlements, significant differences in DDS were found between degrees of IoP, both at street-level and at block-level (Fig. 9).

Firstly, at street-level, the ANOVA test for differences across degrees of IoP was significant (p = 0.003) with a moderate effect size (ω^2 = 0.099). There were no extreme outliers, and no evidence of unequal variances (Levene-test p-value = 0.593). The distributions of sites' DDS-metrics were approximately normal except in IoP 1 (completely spontaneous)–a small group with only 6 sites of which half have sufficient data. The nonparametric ANOVA confirmed the differences between groups (Kruskal test p-value = 0.004, $\eta^2 = 0.1016$). The post-hoc two-sample *t*-test found the strongest evidence of differences between IoP 4 and IoP 2 (p = 0.0006), as well as between IoP 5 and IoP 2 (p = 0.0027).



(caption on next page)

Fig. 7. Results of the significance test. Top (A): Mean DDS across all sites (blue line) compared to simulated null distribution (gray). B1-B4: Four example sites with geotweet distribution. Tweets are clustered for visualisation, with the number of tweets in a cluster indicated as white numbers. The curve below each map displays the simulated reference distribution (gray) relative to the measured disparity (blue line), with values to the left indicating higher density in newer built-up areas and values to the right indicating higher density in older built-up areas. Numbers above the curve indicate left-tailed, and right-tailed p-value respectively. Middle (C): Distribution of DDS values across study sites. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

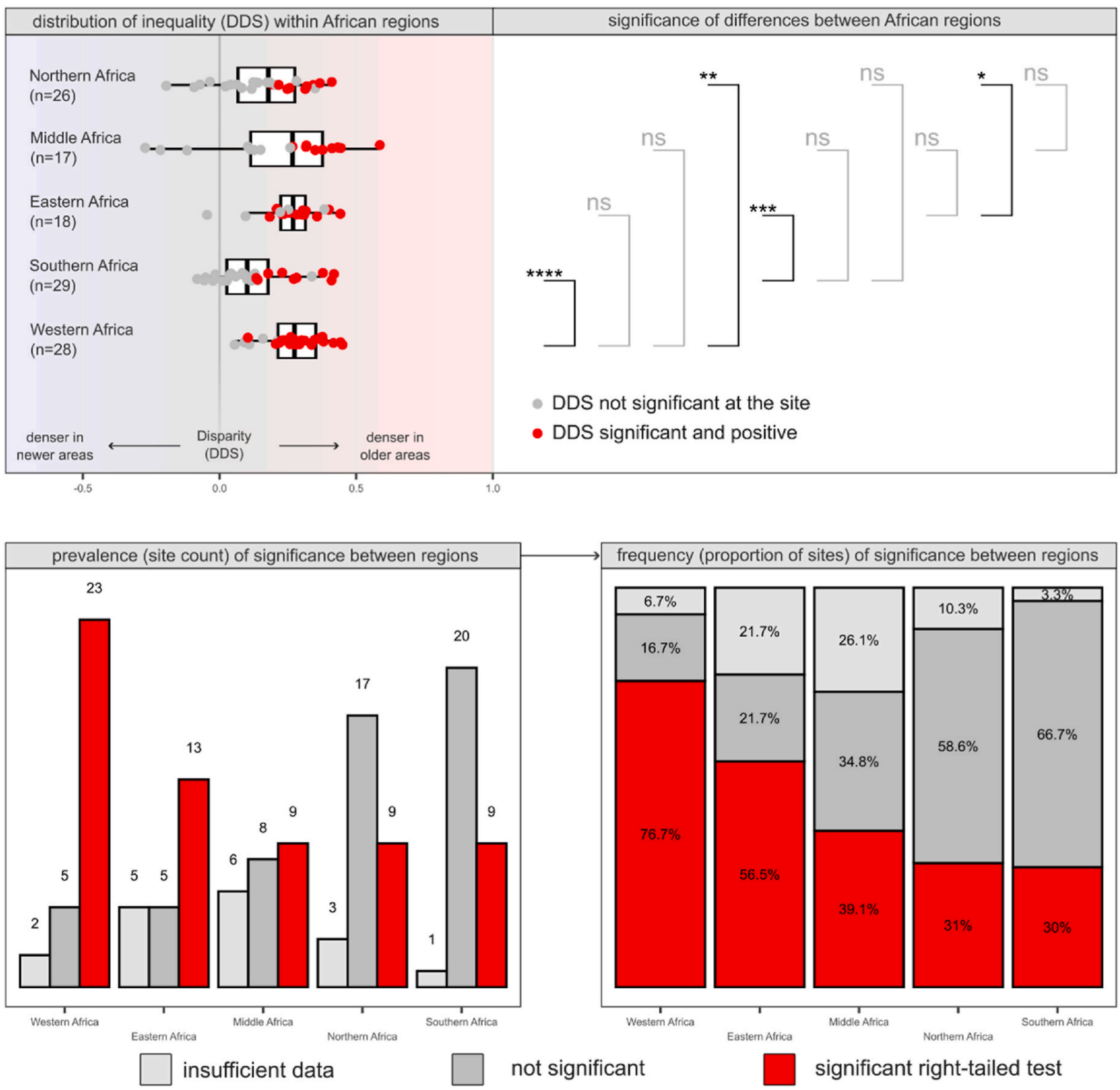


Fig. 8. Comparison of digital disparity (DDS) by UN regions. Top: Comparison of intensity via parametric ANOVA. DDS between regions was assessed with a *t*-test and are encoded as ns: no significant difference; *p $<$ 0.05; **p $<$ 0.01; ***p $<$ 0.001; ****p $<$ 0.0001. Bottom: Prevalence (number of sites) and frequency (proportion of sites) of significant tests within each group.

The frequency of disparity varied across street-level IoP degrees. While for the predominantly spontaneous sites (IoP 2), 16 out of 24 sites (~67 %) showed significant disparity, only 9 out of 28 (~32 %) of the more planned sites with IoP 4–5 did.

At block-level, results were similar. The ANOVA of DDS by degrees of IoP was significant ($p = 0.002$) with a medium effect size ($\omega^2 = 0.109$). There was no evidence of unequal variances (Levene-test p-value $=$ 0.607), but IoP 2 had one extreme outlier (Zango, Angola, in IoP 2) and evidence of non-normality. Hence, we refer to the nonparametric ANOVA which found similar differences between groups (Kruskal test p-value $= 0.001$, $\eta 2$ was moderate at 0.1354). In the post-hoc two-sample *t*-test, the most significant differences were found between IoP 5 and IoP 2 ($p = 0.00096$), although differences also existed between IoPs 4 and 2 ($p = 0.01670$), IoPs 3 and 2 ($p = 0.01145$), IoPs 3 and 5 ($p = 0.02260$), and IoPs 2 and 1 ($p = 0.03601$), although the latter should be interpreted with caution because of the small size of group 1, of which only 4 sites had sufficient data.

Just like at street level, the frequency of disparity varied across block levels (Fig. 10). While for the predominantly spontaneous sites (IoP 2) almost three quarters showed significant disparity, only 1 out of the 7 highly planned sites did. Altogether, the tests strongly indicate that digital disparity differs across IoP degrees.

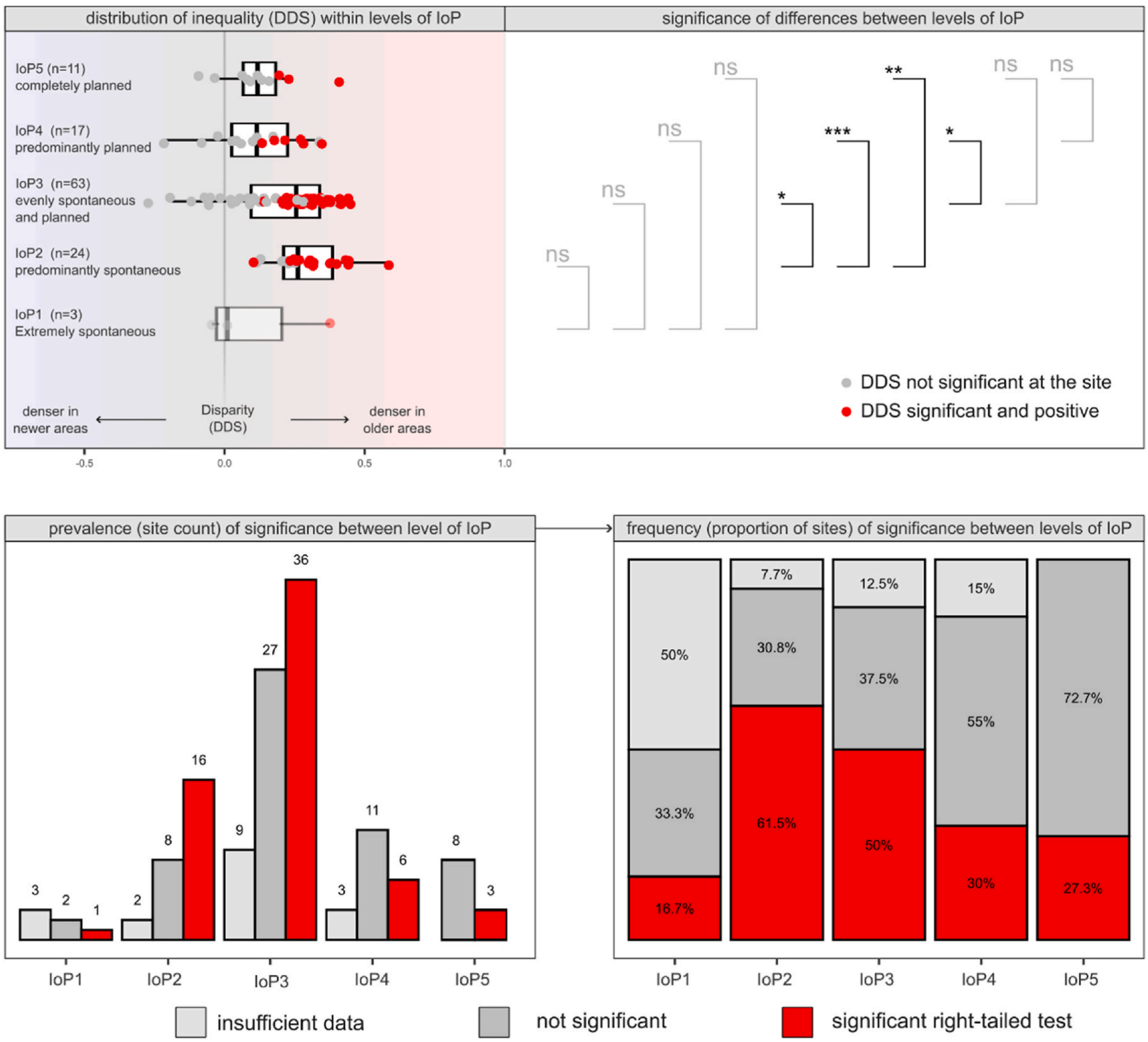


Fig. 9. Comparison of disparity by degrees of intensity of plannedness (IoP) at street-level. Top: Comparison of intensity via the parametric ANOVA test. Differences in disparity (DDS) between degrees of IoP was assessed with a *t*-test and are encoded as: ns = no significant difference; $*p < 0.05$; $**p < 0.01$; $***p < 0.001$; $****p < 0.0001$. Bottom: Prevalence (number of sites) and frequency (fraction of sites) of significant tests within each group.

4.4. *Disparity's relationship with the spatial structure of settlements*

Out of the fourteen models between metrics and DDS, eleven were significant at the 95 % confidence level (Fig. 11), indicating that the digital disparity is linked to aspects of the built-up spatial structure. However, the significant models had limited explanatory power ($R^2 = 0.09$–0.20), suggesting that other factors also influence DDS. Focusing on sites with significant disparity, only 9 out of the 14 models were significant ($R^2 = 0.04$–0.22).

Notably, the spatial metrics Urban Density, Porosity Index, and SPI (>2 km) showed no significant relationship with DDS at the 95 % level. This indicates that built-up density and connectivity are not directly related to the digital density disparity. In contrast, DDS demonstrated a strong correlation with Area-Weighted Standard Distance ($R^2 = 0.17$, $p < 0.01$), Dispersion Index ($R^2 = 0.20$, $p < 0.01$), Urban Compactness ($R^2 = 0.18$, $p < 0.01$), Object Density ($R = 0.09$, $p < 0.01$), Mean Patch Size ($R^2 = 0.09$, $p < 0.01$), Effective Mesh Size ($R^2 = 0.07$, $p < 0.01$), and the SPI at <100 m ($R^2 = 0.12$, $p < 0.01$), 100–500 m ($R^2 = 0.14$, $p < 0.01$), 500–2000 m ($R^2 = 0.12$, $p < 0.01$). This suggests that a higher degree of built-up centrality, compactness, and larger and thus less contiguous patches are associated with higher values of DDS, while a higher degree of disaggregation, dispersion, and more numerous smaller built-up areas are associated with lower DDS.

The model resulting from the stepwise multiple linear regression procedure (Fig. 12 and Table 1) showed that the metrics DI, Euclidean Nearest Neighbor-Distance, and Urban Compactness together explained around a quarter of the variance in disparity ($p < 0.001$, multiple $R^2 = 0.2911$, adjusted $R^2 = 0.2724$). When only including significant sites, the only significant metric was Area-Weighted Standard Distance. The model performed better than single metric models, suggesting that a combination of complementary metrics has a higher association with DDS.

This set of metrics suggests that more compact settlements (high Urban Compactness and SPI, low Dispersion Index) with greater distance between built-up clusters (high Euclidean Nearest Neighbor Distance) and less open spaces (low Urban Density) are associated with greater disparity in digital density. Overall, an urban form which is linked to high digital disparity is characterized by a large and dense compact urban core, with several isolated settlements in the immediate periphery, but far from each other.

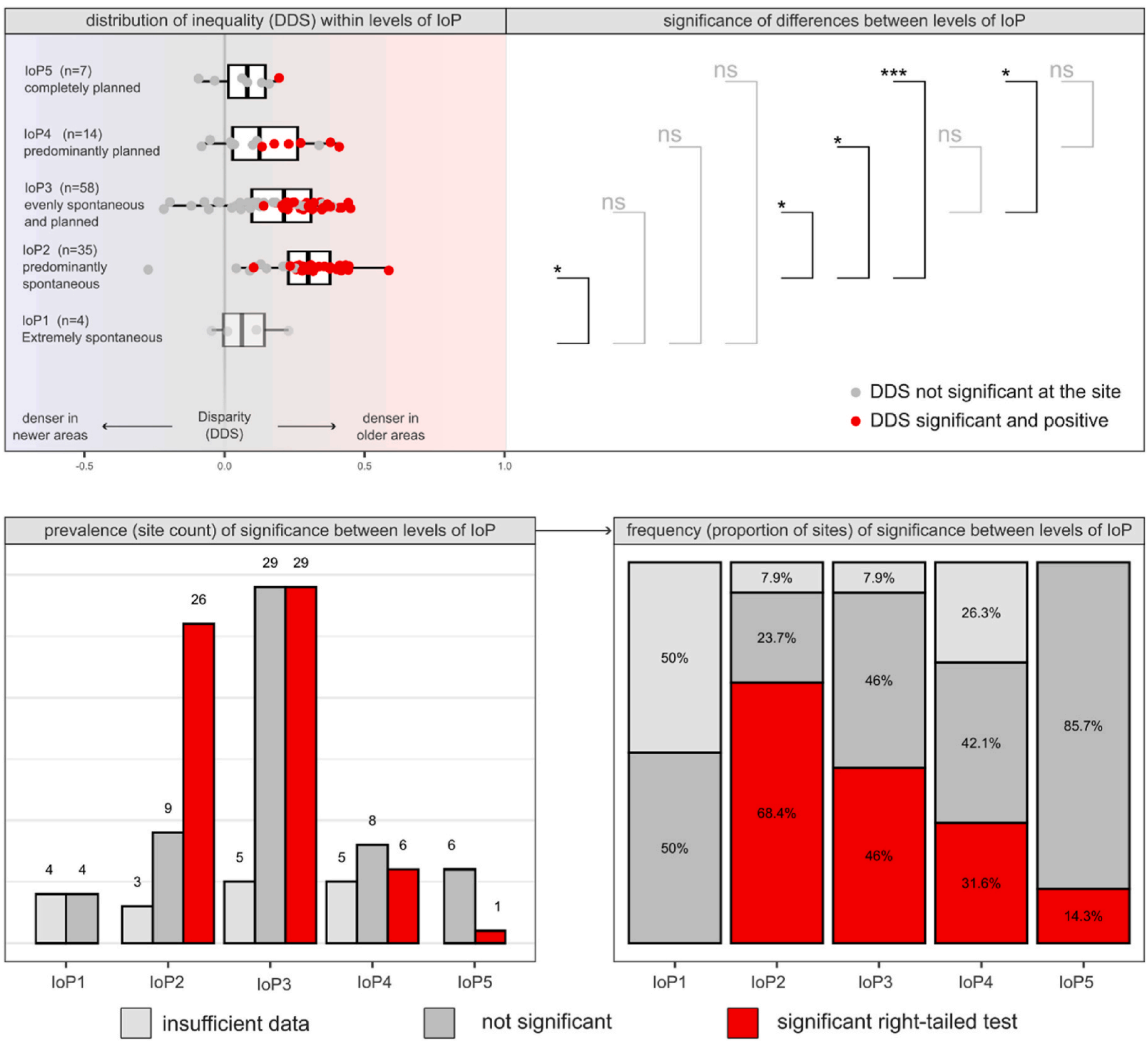


Fig. 10. Comparison of disparity by degrees of intensity of plannedness (IoP) at block-level. Top: Comparison of intensity via ANOVA. Differences in disparity (DDS) between degrees of IoP was assessed with a *t*-test and are encoded as: ns: no significant difference; *p < 0.05; **p < 0.01; ***p < 0.001; ****p < 0.0001. Bottom: Prevalence (number of sites) and frequency (fraction of sites) of significant tests within each group.

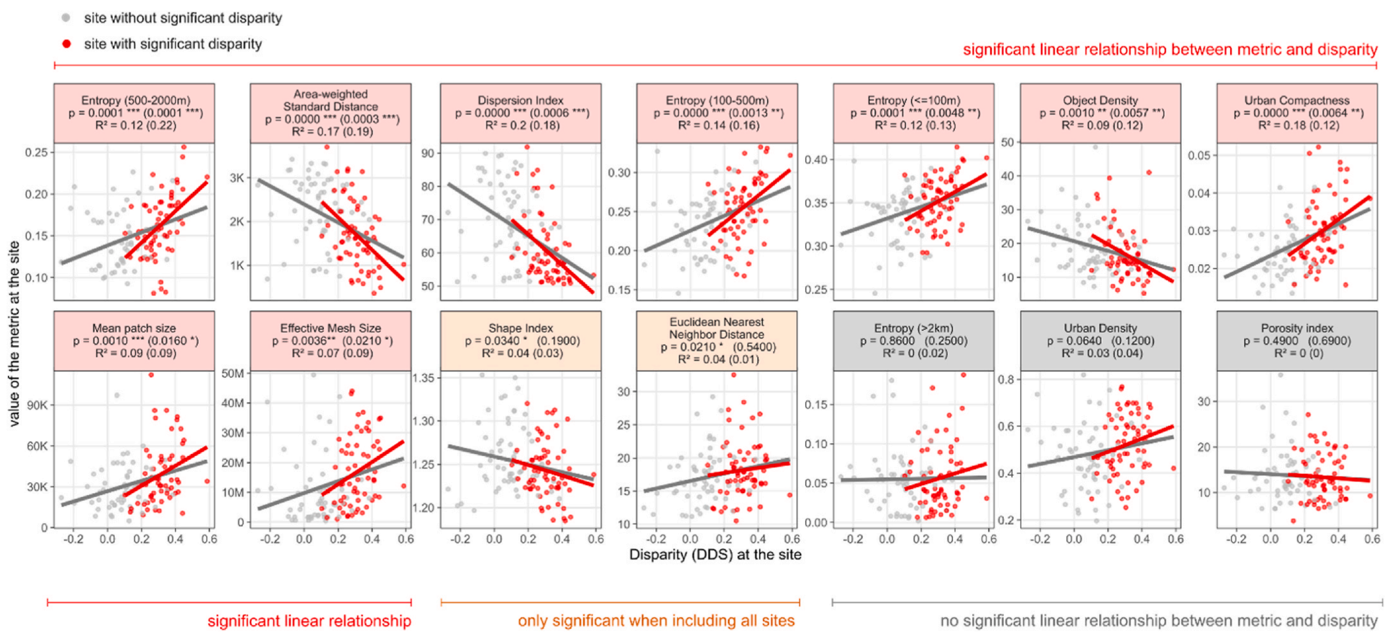


Fig. 11. Relationship between Digital Density Skew (DDS) and spatial metrics of settlement structure, reported by the coefficient of determination (R^2) and p-value (p) of their linear regression. Significance is encoded as $\times$ $p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$. Values in parentheses refer to models that only include significant DDS as observations.

Table 1
Coefficients of the multiple linear regression between DDS and spatial metrics. Significance encodes the p-value as: $^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$. Multiple $R^2 = 0.2911$, Adjusted $R^2 = 0.2724$.

term	coefficient	std. error	t value	p-value (significance)
Intercept	*0.296*	*0.113*	*2.626*	*0.0098 ***
Dispersion Index	−0.005	0.001	−4.077	0.0001 ***
Euclidean Nearest Neighbor Distance	0.007	0.003	2.218	0.0290 *
Urban Compactness	4.000	1.845	2.168	0.0320 *

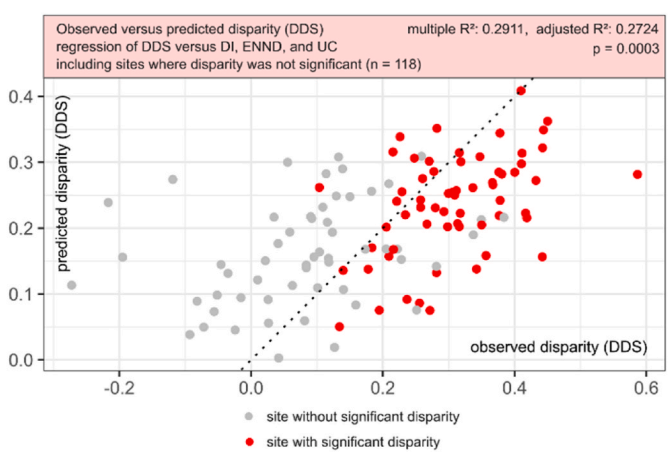


Fig. 12. Observed versus predicted DDS using a multiple linear regression.

5. Discussion

This analysis of geotweet distribution in old and new built-up areas across Africa revealed a hitherto undocumented form of multi-faceted urban digital divide. First, in many of the fastest expanding sites across Africa, we found that digital representation increases with built-up age, at least insofar as measured by geotweet density. Second, this disparity varies in intensity across geographic regions, being strongest in West Africa and weakest in Northern and Southern Africa. Third, this urban digital divide is related to plannedness of the urban structure, with a higher intensity of plannedness linking to lower disparity. And fourth, it is related to the two-dimensional structure of the settlement, particularly built-up fragmentation.

The *first observation*, that a significant disparity exists in many study sites, is the most robust and has two main implications: From a data perspective, it indicates that geotweets, and perhaps GSM in general, are a scarcer data source in new neighborhoods compared to old ones. This does not discourage the use of GSM data altogether, as for 63 of our 135 sites, we find more than 10 K geolocated tweets in the areas that were built-up since 1985. Clearly, data scarcity is not a general rule. Therefore, our conclusion is not that NBAs are digital deserts (Taubenböck et al., 2018), but A) that NBAs may be underrepresented in urban studies relying on GSM; B) that data availability should not be taken for granted when designing studies based on GSM; and C) that biases resulting from unequal data coverage should be considered.

Secondly, our findings can be seen as indication of a digital divide between new and older neighborhoods of African cities. In other words, new neighborhoods are less represented online than older neighborhoods in the same city. There is no single reason for this, as many factors can play a role, from socioeconomic to technical. Based on inspection of our study sites, we assume major influences by density of activity, population, and infrastructure. Unfortunately, there is no sufficiently fine-grained data capturing population density and infrastructure development at Africa-wide scale. Thus, we cannot assess these factors in our study. Even if data were available, the large range of potential factors and the diversity of urban spaces in Africa would require a larger sample than 135 sites to analyze. Our results, however, provide some initial indications which can guide future research: Disparity differs between regions, being greater and more frequent in the study sites of Western Africa than in other parts of Africa, and least frequent in Northern and Southern Africa. Possibly, the higher penetration of internet and communication technologies (Nchake & Shuaibu, 2022) and the slower pace of urbanization in Northern and Southern Africa (UN DESA 2019) have a mitigating effect on the disparity. There is likely some link to the intensity of plannedness, as the disparity is lower and less frequent in more intensely planned sites, which are more common in

Northern and Southern Africa. While our approach is insufficient to determine causalities in the relationship, it is possible that the uniformity in design simplifies the provision of infrastructural resources across the settlement. Further, we find that settlement structure is likely to play a role in the magnitude of the digital divide, although the models' explanatory power is limited. Our interpretation is that there are other (confounding) factors linked to plannedness and settlement structure which influence the magnitude of digital disparity. Almost certainly, the availability of infrastructure and the density of population and cultural or financial capital play a role. Given sufficient data, statistical approaches could disentangle the causal relationships between the physical form of settlements, their built-up age, their socioeconomic structure, and their visibility in the digital sphere. As our study shows, time-series of the settlements' physical form and digital activity already exist at an Africa-wide scale and intra-urban resolution. Unfortunately, it is unlikely that socioeconomic data can be gathered consistently at this scale. Instead, interview-based surveys remain relevant for the study of digital divides. Existing surveys could be expanded to inquire about barriers to social media use and preferences if they do not already do so. Additionally, the inclusion of data on urban functions, land cover, and land use is a promising next step. Such data have been used to assess the impact of local urban functions on digital urban vibrancy (e.g., Lang et al., 2022). The combination of remotely-sensed settlement age with land cover and land use classifications has revealed that the land cover of recently urbanized area differs substantially from older urban areas and between geographic regions (Taubenböck et al., 2025). It will be worth investigating whether the divide between the digital and physical frontier on the African continent is caused by such differences in urban function and land use between newer and older urban areas.

While our big-data approach allowed a comparative study at continent-spanning scale, we acknowledge its limitations. The urban expansion maps are not perfectly accurate and do not distinguish built-up ages earlier than 1985. Data quality varies between regions, potentially affecting cross-region comparisons of DDS. Hence, we implemented a permutation-based test which provides robust estimates of significance. The precision of GPS devices used for the precise geotagging is imperfect, with most studies finding horizontal errors of 5–20 m, but such minor spatial errors lead only to minor errors in the determination of the construction year, and a minor shift in DDS (see Supplement C). More concerning are cases where the number of geotweets and contributors for a site are low. This exacerbates the well-documented demographic bias of Twitter, where a vocal minority contributes most of the tweets (Lemoine-Rodríguez, Biewer, & Taubenböck, 2024; Wojcik & Hughes, 2019). We mitigate the impact of highly active accounts through a weighting procedure and confirm that their behavior is not substantially different from less active accounts for the purposes of this study (see Supplement D). Nevertheless, we acknowledge that geotweet density does not represent the entire digital skin of the urban areas but must be interpreted as a proxy for the online visibility on a particular platform. The choice of Twitter is sensible for our study because it is public and offers geotag functionality. However, in terms of market share, Facebook was found to be the leading social media platform in Africa in 2022, and there are substantial regional variations in the use of platforms and penetration of web technologies in general (Bhanye et al., 2023). Consequently, researchers should choose the platform(s) to fit the geographic setting and goals of the study based on data access. If precise coordinates are not required, geoparsing approaches (Hu et al., 2024) can serve as an alternative to the use of geotags, and provide alternative perspectives. Additionally, the access modes and capabilities of Twitter's API were substantially changed following the rebranding of the platform as "X". As a consequence, the geotweet datasets used in our study can now only be reproduced imperfectly and at substantial cost (Davidson et al., 2023), highlighting the need to diversify web data sources in geographic research.

And just as one platform does not represent social media use as a whole, our selected study sites are not representative of all settlements. Rather, they exemplify the class of fast-growing settlements in Africa where the digital divide we investigated is most acute. Consequently, they are not a representative sample of all growing settlements, and it is possible that patterns might be different in areas of low to medium growth. We also acknowledge that assigning a single IoP or compactness value to each study site is a strong simplification of reality. Each site covers considerable heterogeneity in almost 80 km^2 area and their circular boundary cannot be adapted to morphological subdivisions. Nevertheless, we consider it to be a suitable choice for this study because it is objective and can be derived for any location. Finally, our data is limited in that geotags primarily capture the attention given to a place by local people. While nothing prevents users from commenting on more distant places, the geotagging functionality of Twitter, for most of the study period, only allowed users to tag content with their current location. Consequently, geotagged data includes less content written about places from elsewhere, and thus provides only an incomplete picture of online representation. Geoparsing approaches (Hu et al., 2024) could alleviate this issue by enabling the inclusion of text based on mentions of places. However, most of these approaches rely on existing datasets, such as gazetteers, knowledge-graphs, or training text corpora, which may be biased themselves (Graham and De Sabbata 2020), making them unsuitable for this study.

Our findings are plausible in the African context, and we do not expect them to generalize to the global level because geographic regions differ substantially in their urban history, planning, political systems, and telecommunication infrastructure.

In summary, despite limitations, the converging indications found in the study point strongly towards a first and clear indication of a type of digital urban divide at continental level: A divide between new and consolidated parts of African settlements.

6. Conclusion

By relating maps of urban expansion to GSM posts from Twitter, we find clear evidence of a hitherto undocumented type of digital divide–one that occurs between older and newer parts of the rapidly growing African settlements. Going forward, the perspective provided by GSM posts should be complemented by other information sources on local digitality, and further research with additional socioeconomic data is required to understand the underlying causalities of this digital divide. As a starting point, this study finds that settlements' plannedness and structure are linked to the magnitude of this disparity, and that it should be considered a trend—rather than an inescapable fact—of urban development.

By documenting this disparity, we show that the "urban frontier", which is expanding into the hinterland through the built landscape, often does not run parallel to the "social media frontier", which is lagging behind. It appears that it takes time for these new urban spaces to be filled with places, people, and activities that are representable on the web, and the population with the means and affinity to represent them. Thus, the study reveals a further layer—in the digital sphere—in urbanization processes and contributes to the inclusion of digital perspectives into the broader African Union led 2020–2030 digital transformation strategy, plans and programs, and actions for the transformation and sustainability of Africa's urban spaces.

CRediT authorship contribution statement

Johannes Mast: Writing – review & editing, Writing – original draft, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Marta Sapena:** Writing – review & editing, Software, Methodology, Data curation, Conceptualization. **Henri Debray:** Writing – review & editing, Writing – original draft, Validation, Methodology, Investigation, Data curation, Conceptualization. **Justice Nana Inkoom:** Writing – review & editing, Project administration. **Richard Lemoine-Rodríguez:** Writing – review & editing, Methodology. **Christian Geiß:**

Writing – review & editing, Funding acquisition, Conceptualization. **Hannes Taubenböck:** Writing – review & editing, Supervision, Funding acquisition, Conceptualization.

Funding

This study has been conducted as part of the European Union's Caroline Herschel Framework Partnership Agreement on Copernicus User Uptake (FPCUP) under Grant Agreement No. FPA 275/G/GRO/COPE/17/10042, Action 2019-2-11 "Downstream service/application development for monitoring of environmental indicators", the project MIGRAWARE, Grant No. 01LG 2082C, funded by the German Federal Ministry of Education and Research (BMBF) as part of the programme WASCAL WRAP 2.0, and the research and teaching initiative "A New Focus in English Linguistics: Geolingual Studies", Grant No. 98662, funded by Volkswagen Foundation as part of the programme "Momentum—Funding for Recently Tenured Professors".

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Johannes Mast reports financial support was provided by European Commission. Marta Sapena reports financial support was provided by German Federal Ministry of Education and Research (BMBF). Richard Lemoine-Rodriguez reports financial support was provided by Volkswagen Foundation. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.apgeog.2025.103687.

References

African Union Commission. (2021). *Africa's development dynamics 2021 digital transformation for quality jobs*. OECD Publishing. https://doi.org/10.1787/0a5c9314-en

Altieri, L., & Cocchi, D. (2022). Understanding the expansion of Italian metropolitan areas: A study based on entropy measures. *Environment and Planning B: Urban Analytics and City Science, 49*(2), 447–463. https://doi.org/10.1177/23998083211012699

Altieri, L., Cocchi, D., & Roli, G. (2021). Spatial entropy for biodiversity and environmental data: The R-package SpatEntropy. *Environmental Modelling & Software, 144*(October), Article 105149. https://doi.org/10.1016/j.envsoft.2021.105149

Baddeley, A., & Turner, R. (2005). Spatstat: An R package for analyzing spatial point patterns. *Journal of Statistical Software, 12*(6), 1–42. https://doi.org/10.18637/jss.v012.i06

Barrie, C., & Ho, J. C.-ting (2021). academictwitteR: An R package to access the twitter academic research product Track v2 API endpoint. *Journal of Open Source Software, 6*(62), 3272. https://doi.org/10.21105/joss.03272

Bhanye, J., Shayamunda, R., & Tavirai, R. C. (2023). Social media in the african context: A review study on benefits and pitfalls. In *The palgrave handbook of global social problems* (pp. 1–32). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-68127-2_366-1.

Boy, J. D., & Uitermark, J. (2017). Reassembling the city through instagram. *Transactions of the Institute of British Geographers, 42*(4), 612–624. https://doi.org/10.1111/tran.12185

Cai, J., Huang, B., & Song, Y. (2017). Using multi-source geospatial big data to identify the structure of polycentric cities. *Remote Sensing of Environment, Big Remotely Sensed Data: Tools, Applications and Experiences, 202*(December), 210–221. https://doi.org/10.1016/j.rse.2017.06.039

Debray, H., Kraff, N. J., Xiang Zhu, X., & Taubenböck, H. (2023). Planned, unplanned, or in-between? A concept of the intensity of plannedness and its empirical relation to the built urban landscape across the globe. *Landscape and Urban Planning, 233*(May), Article 104711. https://doi.org/10.1016/j.landurbplan.2023.104711

Goodfellow, T. (2022). *Politics and the urban frontier: Transformation and divergence in late urbanizing east Africa*. Oxford University Press. https://doi.org/10.1093/oso/9780198853107.001.0001

Graham, M., & Sabbata, S. De (2020). The geographic contours of openness. *Making Open Development Inclusive: Lessons from IDRC Research*, 119.

Gumuchian, M.-L. (2012). Pro-government forces seize old gaddafi bastion bani Walid. *Reuters*, 24 October 2012, sec. China https://www.reuters.com/article/world/china/pro-government-forces-seize-old-gaddafi-bastion-bani-walid-idUSBRE89N0WQ/.

Güneralp, B., Lwasa, S., Masundire, H., Parnell, S., & Seto, K. C. (2017). Urbanization in Africa: Challenges and opportunities for conservation. *Environmental Research Letters, 13*(1), Article 015002. https://doi.org/10.1088/1748-9326/aa94fe

Hargittai, E. (2020). Potential biases in big data: Omitted voices on social media. *Social Science Computer Review, 38*(1), 10–24. https://doi.org/10.1177/0894439318788322

Hu, X., Zhou, Z., Li, H., Hu, Y., Gu, F., Kersten, J., Fan, H., & Klan, F. (2024). Location reference recognition from texts: A survey and comparison. *ACM Computing Surveys, 56*(5), 1–37. https://doi.org/10.1145/3625819

Huang, B., & Carley, K. M. (2019). A large-scale empirical study of geotagging behavior on twitter. In *Proceedings of the 2019 IEEE/ACM international conference on advances in social networks analysis and mining* (pp. 365–373). New York, NY, USA: Association for Computing Machinery. https://doi.org/10.1145/3341161.3342870. ASONAM '19.

Jiang, B., Ma, D., Yin, J., & Sandberg, M. (2016). Spatial distribution of city tweets and their densities: City tweets and their densities. *Geographical Analysis, 48*(3), 337–351. https://doi.org/10.1111/gean.12096

Kamana, A. A., Radoine, H., & Nyasulu, C. (2024). Urban challenges and strategies in african cities – a systematic literature review. *City and Environment Interactions, 21*(January), Article 100132. https://doi.org/10.1016/j.cacint.2023.100132

Karami, A., Morgan, L., Webb, F., & Dwivedi, Y. K. (2020). Twitter and research: A systematic literature review through text mining. *IEEE Access, 8*, 67698–67717.

Kruspe, A., Häberle, M., Hoffmann, E. J., Rode-Hasinger, S., Abdulahhad, K., & Zhu, X. X. (2021). Changes in twitter geolocations: Insights and suggestions for future usage. *Proceedings of the 2021 EMNLP workshop W-nut*. https://doi.org/10.48550/arXiv.2108.12251

Lang, W., Lang, H., Hui, E. C. M., Chen, T., Wu, J., & Jahre, M. (2022). Measuring urban vibrancy of neighborhood performance using social media data in Oslo, Norway. *Cities, 131*(December), Article 103908. https://doi.org/10.1016/j.cities.2022.103908

Lemoine-Rodríguez, R., Biewer, C., & Taubenböck, H. (2024). Can social media data help to understand the socio-spatial heterogeneity of the interests and concerns of urban citizens? A twitter data assessment for Mexico city. In H. Carlos-Martinez, R. Tapia-McClung, D. A. Moctezuma-Ochoa, & A. J. Alegre-Mondragón (Eds.), *Recent developments in geospatial information sciences* (pp. 119–133). Cham: Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-61440-8_10.

Lemoine-Rodríguez, R., Mast, J., Mühlbauer, M., Mandery, N., Biewer, C., & Taubenböck, H. (2024). The voices of the displaced: Mobility and twitter conversations of migrants of Ukraine in 2022. *Information Processing & Management, 61*(January), Article 103670. https://doi.org/10.1016/j.ipm.2024.103670

Levene, H. (1960). Robust tests for equality of variances. *Contributions to Probability and Statistics*, 278–292.

Lv, Y., Lan, Z., Kan, C., & Zheng, X. (2021). Polycentric urban development and its determinants in China: A geospatial big data perspective. *Geographical Analysis, 53*(3), 520–542. https://doi.org/10.1111/gean.12236

Malik, M., Lamba, H., Nakos, C., & Pfeffer, J. (2015). Population bias in geotagged tweets. *Proceedings of the International AAAI Conference on Web and Social Media, 9*, 18–27. https://ojs.aaai.org/index.php/ICWSM/article/view/14688.

Marconcini, M., Metz-Marconcini, A., Üreyen, S., Palacios-Lopez, D., Hanke, W., Bachofer, F., Zeidler, J., et al. (2020). Outlining where humans live, the world settlement footprint 2015. *Scientific Data, 7*(1), 242. https://doi.org/10.1038/s41597-020-00580-5

Mast, J., Sapena, M., Mühlbauer, M., Biewer, C., & Taubenböck, H. (2024). The migrant perspective: Measuring migrants' movements and interests using geolocated tweets. *Population, Space and Place, 30*(2), Article e2732. https://doi.org/10.1002/psp.2732

Nchake, M. A., & Shuaibu, M. (2022). Investment in ICT infrastructure and inclusive growth in Africa. *Scientific African, 17*(September), Article e01293. https://doi.org/10.1016/j.sciaf.2022.e01293

OECD. (2025). Africa's urbanisation dynamics 2025: Planning for urban expansion. https://doi.org/10.1787/2a47845c-en.

Petutschnig, A., Clemens, R. H., Resch, B., Krieger, V., & Ferner, C. (2020). Exploratory spatiotemporal language analysis of geo-social network data for identifying movements of refugees. *GI_Forum, 1*, 137–152. https://doi.org/10.1553/giscience2020_01_s137

Rabari, C., & Storper, M. (2015). The digital skin of cities: Urban theory and research in the age of the sensored and metered city, ubiquitous computing and big data. *Cambridge Journal of Regions, Economy and Society, 8*(1), 27–42. https://doi.org/10.1093/cjres/rsu021

Razali, N. M., & Yap, B. W. (2011). Power comparisons of shapiro-wilk, Kolmogorov-smirnov, lilliefors and anderson-darling tests. *Journal of Statistical Modeling and Analytics, 2*(1), 21–33.

Robertson, C., & Feick, R. (2016). Bumps and bruises in the digital skins of cities: Unevenly distributed user-generated content across US urban areas. *Cartography and Geographic Information Science, 43*(4), 283–300. https://doi.org/10.1080/15230406.2015.1088801

Royston, J. P. (1982). An extension of Shapiro and wilk's W test for normality to large samples. *Journal of the Royal Statistical Society. Series C (Applied Statistics), 31*(2), 115–124. https://doi.org/10.2307/2347973

Sanderson, R., Franklin, R., MacKinnon, D., & Matthews, J. (2024). Left out and invisible? : Exploring social media representation of "left behind places". *Geojournal, 89*(1), 37. https://doi.org/10.1007/s10708-024-11010-z

Sapena, M., Mast, J., Schoepfer, E., & Taubenböck, H. (2024). Do-it-yourself built-up mapping tool: A practical cloud-based solution using Sentinel imagery for mapping

urban expansion in Africa. *SSRN Scholarly Paper. Rochester, NY*. https://doi.org/10.2139/ssrn.4762416
Sapena, M., Ruiz, L., & Taubenböck, H. (2020). ‘Analyzing links between spatio-temporal metrics of built-up areas and socio-economic indicators on a semi-global scale’. *ISPRS International Journal of Geo-Information, 9*(7), 436. https://doi.org/10.3390/ijgi9070436
Saunders, D. (2010). *Arrival city: How the largest migration in history is reshaping our world*. Random House.
Şen, İ. (2024). ‘A decolonial analysis of urban and regional planning education in the global south’. http://hdl.handle.net/11527/26452.
Smith, P., Gould, M., & Manley, E. (2025). The rise and fall of twitter data in urban analytics. *Findings*, (March)https://doi.org/10.32866/001c.133245
Steiger, E., Westerholt, R., Resch, B., & Zipf, A. (2015). Twitter as an indicator for whereabouts of people? Correlating twitter with UK census data. *Computers, Environment and Urban Systems, 54*(November), 255–265. https://doi.org/10.1016/j.compenvurbsys.2015.09.007
Taubenböck, H., Debray, H., Qiu, C., Schmitt, M., Wang, Y., & Zhu, X. X. (2020). Seven city types representing morphologic configurations of cities across the globe. *Cities, 105*, Article 102814.
Taubenböck, H., Mast, J., Geiß, C., Wurm, M., Esch, T., & Seto, K. C. (2024). Global differences in urbanization dynamics from 1985 to 2015 and outlook considering IPCC climate scenarios. *Cities, 151*(August), Article 105117. https://doi.org/10.1016/j.cities.2024.105117
Taubenböck, H., Mast, J., Lemoine Rodríguez, R., Debray, H., Wurm, M., & Geiß, C. (2025). Was global urbanization from 1985 to 2015 efficient in terms of land consumption? *Habitat International, 160*(June), Article 103397. https://doi.org/10.1016/j.habitatint.2025.103397
Taubenböck, H., Staab, J., Zhu, X., Geiß, C., Dech, S., & Wurm, M. (2018). Are the poor digitally left behind? Indications of urban divides based on remote sensing and twitter data. *ISPRS International Journal of Geo-Information, 7*(8), 304. https://doi.org/10.3390/ijgi7080304
Taylor, Z. (2013). Rethinking planning culture: A new institutionalist approach. *Town Planning Review, 84*(6), 683–702. https://doi.org/10.3828/tpr.2013.36
Törnberg, P. (2023). The urban attention economy: How our cities are rebuilt for social media. *Inside Media: A Journal of Cities and Culture, 8*(1). https://www.mediapolisjournal.com/2023/03/urban-attention-economy/.
Törnberg, P., & Uitermark, J. (2022). Urban mediatization and planetary gentrification: The rise and fall of a favela across media platforms. *City & Community, 21*(4), 340–361. https://doi.org/10.1177/15356841211068521
Twitter Inc. (2022). Twitter API for academic research | products. https://developer.twitter.com/en/products/twitter-api/academic-research.
UN, D. E. S. A. (2019). *World urbanization prospects: The 2018 revision (ST/ESA/SER.A/420)*. New York: United Nations.
United Nations Statistical Office. (1982). *Standard country or area codes for statistical use, 49*. UN.
Wentrup, R., Ström, P., & Nakamura, H. R. (2016). Digital oases and digital deserts in sub-saharan Africa. *Journal of Science and Technology Policy Management, 7*(1), 77–100. https://doi.org/10.1108/JSTPM-03-2015-0013
Wojcik, S., & Hughes, A. (2019). Sizing up twitter users. *PEW Research Center, 24*, 1–23.
Yin, J., & Guangqing, C. (2021). Characterizing people's daily activity patterns in the urban environment: A mobility network approach with geographic context-aware twitter data. *Annals of the Association of American Geographers, 111*(7), 1967–1987. https://doi.org/10.1080/24694452.2020.1867498
Zhu, L., & Lerman, K. (2016). Attention inequality in social media. https://doi.org/10.48550/arXiv.1601.07200.