EPJ Data Science
a SpringerOpen Journal

REGULAR ARTICLE   **Open Access**

The impact of social segregation on human mobility in developing and industrialized regions

Alexander Amini, Kevin Kung, Chaogui Kang, Stanislav Sobolevsky[*] and Carlo Ratti

[*]Correspondence: stanly@mit.edu
SENSEable City Laboratory,
Massachusetts Institute of
Technology, 77 Massachusetts
Avenue, Cambridge, 02139, USA

Abstract

This study leverages mobile phone data to analyze human mobility patterns in a developing nation, especially in comparison to those of a more industrialized nation. Developing regions, such as the Ivory Coast, are marked by a number of factors that may influence mobility, such as less infrastructural coverage and maturity, less economic resources and stability, and in some cases, more cultural and language-based diversity. By comparing mobile phone data collected from the Ivory Coast to similar data collected in Portugal, we are able to highlight both qualitative and quantitative differences in mobility patterns - such as differences in likelihood to travel, as well as in the time required to travel - that are relevant to consideration on policy, infrastructure, and economic development. Our study illustrates how cultural and linguistic diversity in developing regions (such as Ivory Coast) can present challenges to mobility models that perform well and were conceptualized in less culturally diverse regions. Finally, we address these challenges by proposing novel techniques to assess the strength of borders in a regional partitioning scheme and to quantify the impact of border strength on mobility model accuracy.

Keywords: predictive human mobility; social networks; cultural diversity

Introduction

Transportation and communication networks form the fabric of industrialized nations. The roll-out of such infrastructure in such regions can play a major role in supporting, or deterring, a regions' ability to thrive economically and socially. Likewise, citizens' use of these networks can tell us much about the region, including insight on how ideas and diseases may be spreading, or how to most effectively augment services, such as health care and education [1].

Existing studies of mobile phone data have given us insight on numerous aspects of human mobility [2–6]. However, these studies tend to focus on regions with the highest mobile phone coverage, which also happens to be in more stable, mature, and developed regions. Thus, the models produced based on this data might not be as appropriate for developing regions with a substantially different patterns of social interactions and human mobility. However, these highly industrialized and wealthy regions represent less than one-third of the world's population, with the remaining two-thirds living in developing and less economically mature regions. Accurate models for developing regions are critical

Springer

as these regions are facing the most rapid demographic and economic shifts worldwide, and are in even greater need of such models to help inform policy makers, urban planners, and service providers. Yet, little work has been done to assess the appropriateness of models conceptualized for industrialized regions for use in developing regions.

Obtaining a comprehensive and accurate dataset of the telecommunications activity in developing regions can be extremely difficult due to security and privacy considerations, limited coverage by any single provider, and the need for a rigorous data capture methodology and infrastructure. The Data4Development (D4D) dataset [7] provided a unique opportunity by collecting data throughout the Ivory Coast and releasing it specifically for research purposes, so that developing regions could also be analyzed in greater detail. The contrast of long-standing cultural and linguistic diversity with relatively recent and rapid urbanization offered researchers a unique opportunity to understand the communication and mobility patterns and needs of a developing nation during key phases of its transformation.

Our study brought the D4D data from Ivory Coast together with mobility data from an industrialized nation (Portugal) in order to assess the ability of human mobility models developed for industrialized regions to accurately model developing regions. We focus on the comparison of these two countries at very different stages in their industrialization and in their levels of cultural and linguistic diversity, and sheds new light on the applicability of metrics and models conceptualized for industrialized regions to developing regions. Our results demonstrate the importance of considering cultural and linguistic diversity in the construction of new models to address the challenges of developing regions. The insights gained from our study have important applications to policymaking, urban planning, and the services deployments that are transforming Ivory Coast and many other developing countries.

In the following sections, we provide additional details on the data used in this study, the results derived, and the conclusions drawn.

Related works

Leveraging mobile phone data to elucidate and quantify many aspects of human life is growing in popularity. For example, mobile phone data has been used to gain insights from a diversity of cultures, ranging from university students to professionals in the US, Finland, and Africa [8]. Targeted cultural patterns included pace of life, reaction to outlier events, and social support, as opposed to the mobility focus of our study. Eagle et al used mobile phone communication logs and top up records to conduct a comprehensive comparison of urban and rural life within a small country, as opposed to across countries as targeted by our research [9]. Mobile phone data has also been used to study the seasonal consumption patterns of tourists in Estonia [10].

As part of the D4D competition, researchers studied a wide range of topics ranging from social behaviors, economics, health, transportation, and mobility. Several studies [11–13] considered mobility patterns for the purpose of improving the planning or efficiency of the transportation systems. While some articles [14, 15] targeted mobility within the largest city, Abidjan, and others considered mobility across the country, none tackled the challenge of assessing mobility models created for mature and industrialized nations on the developing nation of Ivory Coast. Additionally, ours is the first study to have considered the linguistic and cultural barriers and affinities that we have shown to be significantly

stronger in the developing nation of Ivory Coast, in comparison the more industrialized nation of Portugal. Thus our study represents a novel and important contribution to understanding the challenges of creating globally applicable mobility models.

Data description

We used four datasets to assess and compare the human mobility patterns in the Ivory Coast and Portugal. The first dataset, *D1* was provided by Orange Telecom as *SET2*, via the Data for Development (D4D) Challenge [7]. This dataset was based on anonymized Call Detail Records (CDRs) of 2.5 billion calls and SMS exchanges between 5 million users December 1, 2011 until April 28, 2012 (150 days).

SET2 contains consecutive call activities of each subscriber over the study period. Each record in this dataset represents a single connection to an antenna and contains the following fields: timestamp, anonymized ID of the user, and the antenna ID they connected to. To further anonymize this data, the original dataset was subsampled to the calls of 50K randomly sampled individuals for each of 2-week periods in the dataset. The geographical positions of the antenna for *D1* was also provided and visualized in Figure 1B according to the density of antennae in every region. Black lines overlaid on the country represent the corresponding first-level administrative boundaries (19 Ivorian régions and 20 Portuguese
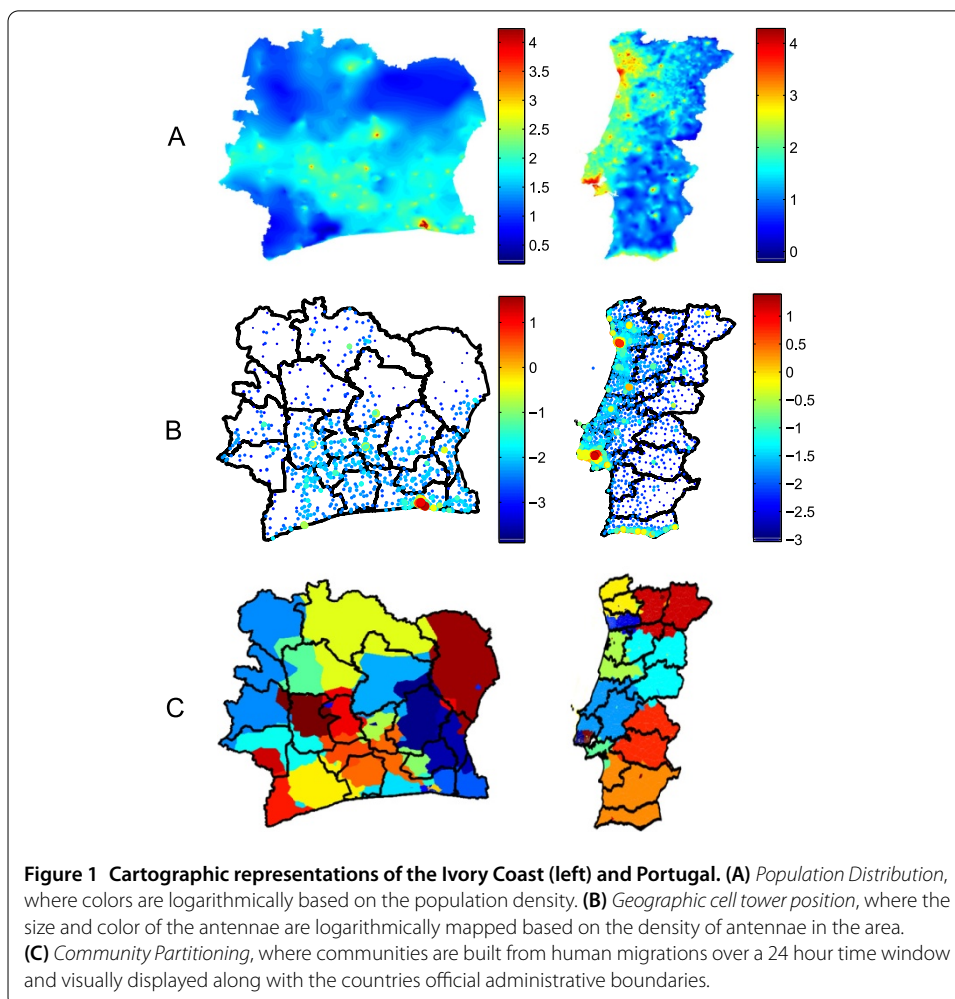

Figure 1 Cartographic representations of the Ivory Coast (left) and Portugal. (A) *Population Distribution*, where colors are logarithmically based on the population density. **(B)** *Geographic cell tower position*, where the size and color of the antennae are logarithmically mapped based on the density of antennae in the area. **(C)** *Community Partitioning*, where communities are built from human migrations over a 24 hour time window and visually displayed along with the countries official administrative boundaries.

districts). Records without antenna IDs were removed; 107 antennae had no calls and 128 antennae had no population movements.

The second dataset, *D2*, provided 400 million anonymized CDRs across Portugal for the time period of January 1, 2006 to December 31, 2007. *D2* was also provided by Orange Telecom with 2000 antennae distributed across Portugal, and the same data fields as in *D1*.

Datasets *D3* and *D4* provided a high-resolution population density data for Ivory Coast [16] and Portugal [17], respectively. To map the population data to the antennae, we created a Voronoi tessellation [18] of each country based on the antennae location. For the 12 locations that had 2-3 antennas in a single location, those 2-3 antennas were collapsed into a single Voronoi cell. Each antenna was assigned the total population within the corresponding Voronoi cell. Figure 1A provides a logarithmic scale population density distribution map using the data from *D3* and *D4*. The population maps were created as an interpolation of the population density at each antenna.

Although used widely for human mobility studies, mobile phone data provides only a proxy for human mobility, for example, callers are tracked only to the spatial resolution of the antenna (which may be up to 70 km, depending on tower height and terrain), and usually only when the phone is in use while not everyone uses a mobile phone while traveling. However, even in developing regions, mobile phone penetration is high. Ivory Coast has 85% mobile phone coverage, with Orange Telecom (the provider of the data for Ivory Coast and Portugal used in this study) being the top mobile phone provider having a market share of 42.5%. In general mobile phone penetration in Portugal is almost absolute, while the total number of phone accounts is even higher than the total number of people - 142% of the country population owns a mobile device, with Orange Telecom at 19% of market share.

Collective mobility patterns

We first performed a bulk mobility pattern analysis based on *D1* and *D2* by plotting the probability density function $P(\Delta r)$ of the individual travel distances (or jump sizes) Δr in a trace of agglomerated de-identified callers over a period of two weeks, for Ivory Coast (left plot, solid black line) and Portugal (right plot, solid black line), as shown in Figure 2A. The distributions were qualitatively similar to each other except that at the administrative level, the distributions in Ivory Coast are much more scattered than those observed in Portugal, suggesting greater regional variance. We fit the density function to a truncated power law of the form $P(\Delta r) = (\Delta r + \Delta r_0)^{-\beta} \exp(-\Delta r/\kappa)$, as described in [2], where Δr_0, β, and κ are the fit constants. While the two distributions had similar cutoff distance ($\kappa_{\text{Portugal}} = 106 \pm 10$ km; $\kappa_{\text{Ivory}} = 122 \pm 5$ km), the two countries have slightly different power law coefficients ($\beta_{\text{Portugal}} = 1.37 \pm 0.06$; $\beta_{\text{Ivory}} = 1.62 \pm 0.03$). This higher coefficient in Ivory Coast indicates that the likelihood of displacement generally decays faster with distance in comparison to Portugal.

We also investigated regional differences in the mobility patterns. In both cases of Ivory Coast and Portugal, we identified the first level administrative boundaries as the highest country-defined level of partitioning. For Ivory Coast these are called 'régions,' while in Portugal they are referred to as 'districts.' We partitioned the mobility data by the different level-one administrative regions and overlaid the same density functions specific for each administrative region on the same plots above. Different administrative regions are
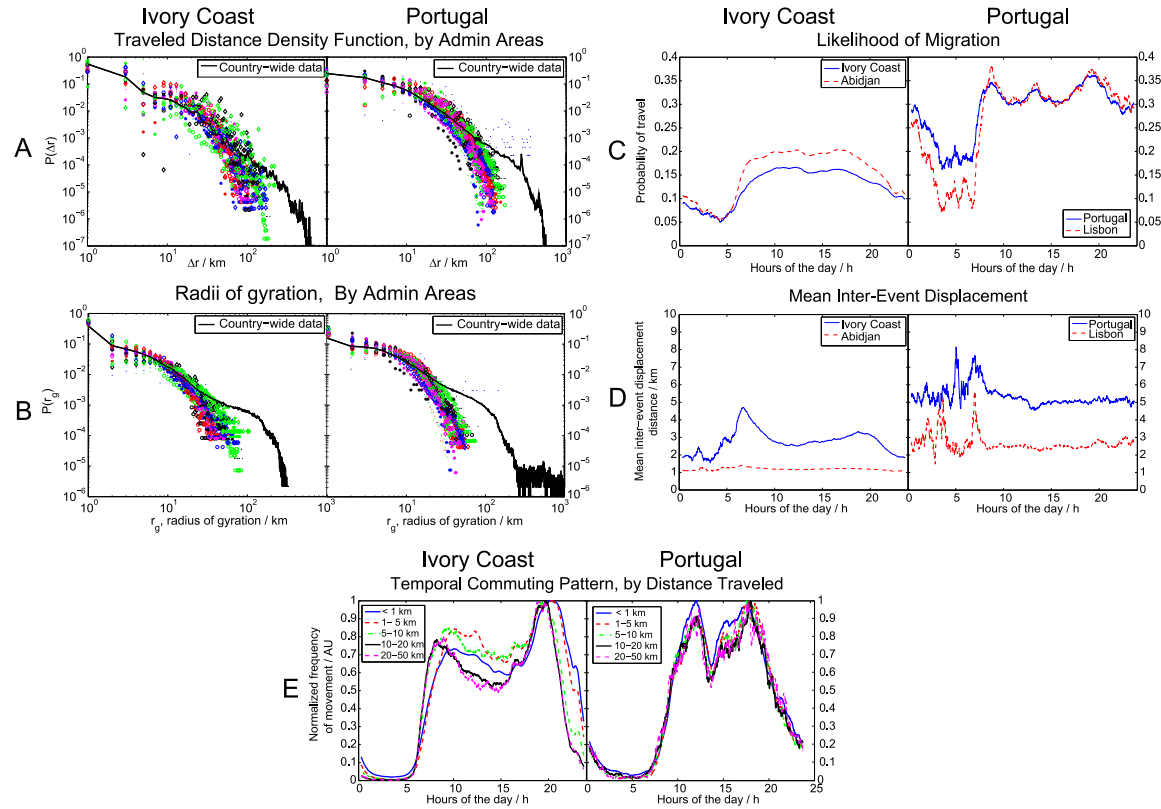
Figure 2 Comparative mobility functions for Ivory Coast (left) and Portugal (right). Probability density functions for *distance traveled* **(A)** and *radius of gyration* **(B)** for each administrative area (differentiated by different color/shape markers). Comparison of country-wide data to capital city daily commuting profiles through respective *probabilities of displacement* **(C)** and *mean inter-event migration distance* **(D)**. Country wide temporal commuting profiles *separated by the distance traveled* **(E)**.

identified by different scatter marker types and colors. We observed the same truncated power law behavior across the different regions, but the Ivory Coast regions exhibited significantly greater diversity than similarly defined regions in Portugal. This would indicate that in Ivory Coast the likelihood that people migrate and commute with respect to distance is much more dependent on what part of the country they are in, as opposed to in Portugal where the different administrative regions show very little diversity from each other.

Another important metric for assessing mobility patterns is the radius of gyration. As defined in [2], the radius of gyration for each caller is the characteristic distance traveled by each caller when observed up to time t, and is computed as the probability distribution function of the mean squared variance of the center of mass of each user's set of catchment locations. The results are plotted in Figure 2B, with t = the period of data collection for each of the datasets.

The distributions plotted in Figure 2B and show that the bulk mobility data from Ivory Coast adheres well to the scale-free framework proposed in [2]. The similarity in the bulk mobility characteristics between Ivory Coast and Portugal serves to strengthen the argument that we can make valid comparisons between the two datasets, as described in the sections below.

Commuting patterns

Daily commuting patterns are a critical component of any region's mobility requirements. Displacement is defined as movement from one cell tower to another cell tower between two consecutive calls, and is a key marker for assessing mobility. To focus on daily commuting patterns, we excluded data collected during weekends, and computed the fraction of inter-call events that were accompanied by displacements in a moving 40-minute window of time for Ivory Coast and Portugal. We averaged the fraction of displacement for each 40-minute window across 45 weekdays to get a 24-hour temporal profile of the probability of displacement during a workday.

The first and probably the most significant difference is the absolute difference in the probability of displacement, which can be seen in Figure 2C. We observe that in Portugal, in a given period, people are much more mobile compared to their counterparts in Ivory Coast.

Both countries exhibit a commuting pattern; there is a sharp rise in the probability of displacement around 7-9 a.m. The evening decline is not as sharp, suggesting that people leave work at different times in the evening.

Significant quantitative differences between the countries can also be seen throughout the day. In Portugal, people in Lisbon and across the nation exhibited similar likelihood to commute during the busiest hours. However, a significantly higher percentage of people in Abidjan were mobile than across the nation. Additionally, while displacement levels in Abidjan and across Ivory Coast were similar during the lowest period (4-7 a.m.). Displacement for the same period is significantly higher for Portugal than for Lisbon, and is likely an indicator of more significant numbers of suburban commuters in Portugal than in Ivory Coast.

Figure 2D provides a comparison of the mean migration distances between the 2 countries for the same period. Here again, the average distance traveled is significantly less in Ivory Coast and its capital city, than in Portugal. In the country-wide data, we observe a

sharp increase in the mean inter-event displacement distance near the morning peak commute (around 5-9 a.m.) in both Ivory Coast and Portugal. However, the spike in distance encountered in Lisbon during morning commute does not occur in Abidjan. This difference may be indicative of people both living and working in close proximity in Abidjan, as opposed to commuting in from outside or across the city as is often the case in developed regions with more comprehensive public transport facilities.

We examined the country-specific commuting pattern more closely by looking at how the distance commuted may affect the daily behaviors. The observed distances traveled were binned (0-1 km, 1-5 km, 5-10 km, 10-20 km, and 20-50 km), and the daily temporal profile of the probability of displacement was computed for each bin for the two countries, as shown in Figure 2E.

The temporal profiles for both Ivory Coast and Portugal show a bimodal pattern in Figure 2E. For Ivory Coast the morning peak is around 7 a.m. and the evening peak is around 8 p.m., as opposed to roughly 10 a.m. and 6 p.m. for Portugal. Note that the peaks for Portugal are much sharper than those of Ivory Coast. Overall these differences indicate a shorter prime commuting period for Portugal, which is likely indicative of the ability for commuters to travel more efficiently to their destinations.

Community structure

Large networks, such as the telecommunications or transportation networks of a nation, often exhibit community structure, i.e., the organization of vertices into clusters with many edges joining vertices of the same cluster and comparatively few edges joining vertices of different clusters. Identifying the community structure in such networks has many applications, such as better placement and provisioning of services. Recently, this type of community structure analysis has been performed on land-line communications in Great Britain [19], mobile connections in Belgium [20], United States [21], and various other countries across Europe, Asia, and Africa [22, 23]. While there is research investigating the impact of physical human mobility on the space-independent community structure [21] there has been a lack of similar research for developing nations.

Network modularity [24] is a measure of the strength of the division of a network into clusters. Networks with high modularity have dense connections between nodes within clusters, and sparse connections between nodes in different clusters. Modularity is computed as the fraction of edges that fall within a cluster, minus the expected such fraction if the edges were distributed at random with respect to the node strength distribution. The value of modularity lies in the range $[-1, 1]$, and is positive if the edges within groups exceeds the number expected on the basis of chance.

The definition of network modularity is:

$$Q = \frac{1}{2m} \sum_{ij} \left[A_{ij} - \frac{k_i k_j}{2m} \right] \delta(c_i, c_j), \qquad (1)$$

where A_{ij} is the weight of the link from i to j, k_i is the sum of the weights from node i, c_i is the community that node i was assigned to, $m = \frac{1}{2} \sum_{ij} A_{ij}$, and $\delta(c_i, c_j)$ is 1 if $c_i = c_j$ and 0 otherwise.

High modularity in mobility networks may point to an efficient organization of residences, employment, and services all in close proximity, or it may point to restrictive policies or infrastructures that limit free movement across communities. We were interested

Table 1 Comparison of different similarity indices to compare similarity between community partitioning (generated from network of human mobility) and the respective administrative boundaries

Similarity Index	Portugal	Ivory Coast	Difference
Wallace [30]	0.482813	0.199134	0.283679
Adjusted Rand [31]	0.495536	0.258351	0.237184
Jaccard [32]	0.377999	0.184733	0.193266
Fowlkes-Mallows [33]	0.553788	0.378296	0.175491
Melia-Heckerman [34]	0.659385	0.515347	0.144037
Hubert [35]	0.806411	0.707795	0.098615
Larsen [36]	0.58273	0.525396	0.057334
Rand [31]	0.903205	0.853898	0.049308

in the community structure of developing nations, such as Ivory Coast, especially in comparison to more developed nations, such as Portugal.

We used datasets *D1* and *D2* to build the human mobility networks and identify the community structure of antennae within the Ivory Coast and Portugal. We set nodes to the locations of each cell tower, and edges to the total number of migrations of all people that placed two consecutive calls between the two nodes within a time window of 24 hours. We tested the following Community Detection algorithms: Louvain [25], Le Martelot [26], Newman [27], Infomap [28], and a new method of community detection suggested in [29]. We computed the modularity of community structures identified by each of these methods. The method described in [29] provided the highest modularity, and was subsequently chosen to be used for this part of the study. Figure 1C graphically compares the communities identified (in color) with their first level administrative boundaries (outlined in black).

An especially interesting difference in the communities identified for Ivory Coast and those identified for Portugal was the similarity between identified communities and the official administrative boundaries of the nations. We calculated 7 different clustering coefficients, each representing the qualitative similarity between the two different partitions of each region. While the communities identified for Portugal exhibited high similarity with the 20 official administrative boundaries (districts), this was not the case for the Ivory Coast's 19 official administrative boundaries (régions). As shown in Table 1, communities identified for Portugal show significantly higher similarity (as much as 28% higher clustering coefficients) to administrative boundaries, in comparison to that of the Ivory Coast.

While this significant difference in community and official boundary alignments may be attributable to the layout of infrastructure along official boundaries, we began to question whether there might be more fundamental differences. Previous studies have shown that other factors, such as geographical features, can play an important role in how communities are formed and services are sought [19, 37]. However, little has been done to investigate the direct impact of culture and language on human mobility.

Ivory Coast represents an especially interesting context to investigate cultural and linguistic influences on mobility within a single nation. The Ivory Coast is a nation made up of more than 60 distinct tribes, classified into 5 principle regions [38]. The official language is French, although many of the local languages are widely used, including Baoulré, Dioula, Dan, Anyin and Cebaara Senufo, and an estimated 65 languages are spoken in the country.

Intuitively, these cultural and linguistic differences are likely to influence mobility patterns in the region. However, it is also known that as regions become more industrialized, cultural ways are often blended or lost altogether. Portugal represents an interesting con-

text for the latter, as Portuguese is the single national language of Portugal, and any tribal boundaries pre-date Roman times.

Due to the vast differences seen in the network community structure to administrative-defined boundaries, our goal in the next section was to understand if tribal structure of the Ivory Coast could be the influential factor of the mobility patterns in the region.

Tribal community analysis

Since the communities detected in Section IV exhibited low similarity to administrative boundaries, we began to investigate the impact of certain factors present in the Ivory Coast that may be attributable to such a substantial difference. Since people and their behaviors throughout a large portion of Africa are still impacted by their tribal affiliations while any tribal boundaries in Portugal predate Roman times, studying the tribal boundary impact in Ivory Coast represented a perfect example of this. We generated digital shape files for each of the eight distinct tribal regions in Ivory Coast [39]. Such tribal maps do have a level of uncertainty associated with them due to migrations over time; however, we used the most recent versions of tribal boundaries available.

We then modified the community detection approach to use the tribal boundaries as the Level 1 boundaries and subsequently ran a hierarchical community detection using the Louvain method [25] in each of these Level 1 tribal partitions in order to produce the sub-communities inside different tribal regions which we will refer to as sub-tribal communities. The Louvain method provided the closest final number of partitions to the administratively defined subprefectures of the algorithms that we tested, and was thus chosen as the most appropriate for current purposes. By doing so, we were able to generate sub-tribal communities while also conserving the physical shape of each tribal region. Figure 3A demonstrates this by showing the official prefecture and subprefecture boundaries (left) and the tribal and sub-tribal communities (right) that were created using this approach. The larger first level partitions (prefecture and tribal) are indicated as a single color, while the smaller secondary partitions (subprefecture and sub-tribal) are indicated as black lines within their respective first level partition. The number of subprefectures and sub-tribal regions are very similar (approximately 250 and 220 respectively) and thus can be used as a valid comparison against each other.

As a first measure of impact of tribes on mobility, we aggregated the mobility network between communities. Figure 3B provides a plot of the mobility network with each node representing a sub-tribal community, and each edge is colored on a logarithmic scale to reflect the number of migrations between the connected nodes in the mobility network. Nodes are colored with the same logarithmic scheme and represent the sum of all migrations into this location. The intra-tribal community mobility network is plotted separately from the inter-tribal community mobility network, in order to facilitate comparison. The intra-tribal network plots only those edges between communities in the same tribe. The inter-tribal network plots only edges between communities of differing tribes.

This diagram provides a first insight into tribal influences on mobility. Note that the number of intra-tribal migrations (as indicated by the color coding of edges) dwarfs the number of inter-tribal migrations. Additionally, the inter-tribal migrations are largely dominated by connections to the largest city, Abidjan.

The impact of the tribal borders versus administrative borders in Ivory Coast on human mobility could be also directly demonstrated by measuring the strength of normalized mobility fluxes at different geographical distances in case they cross or do not cross those
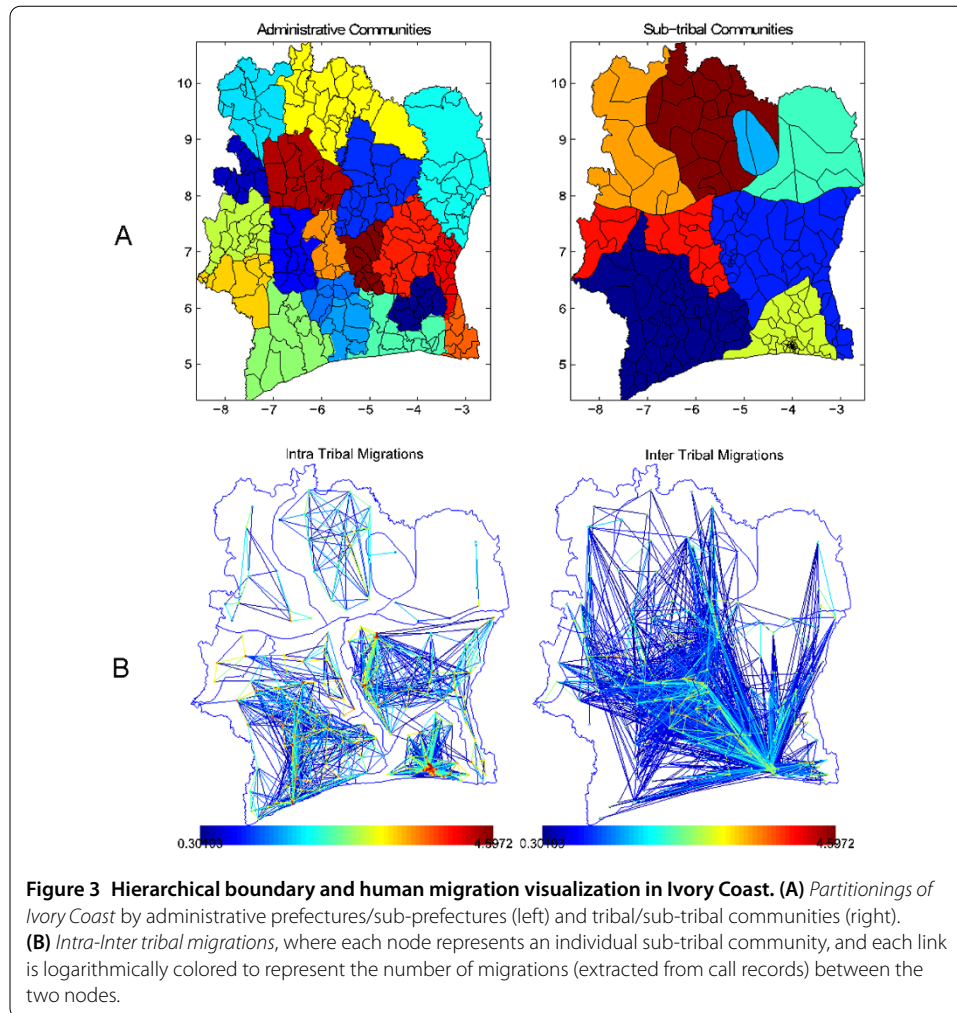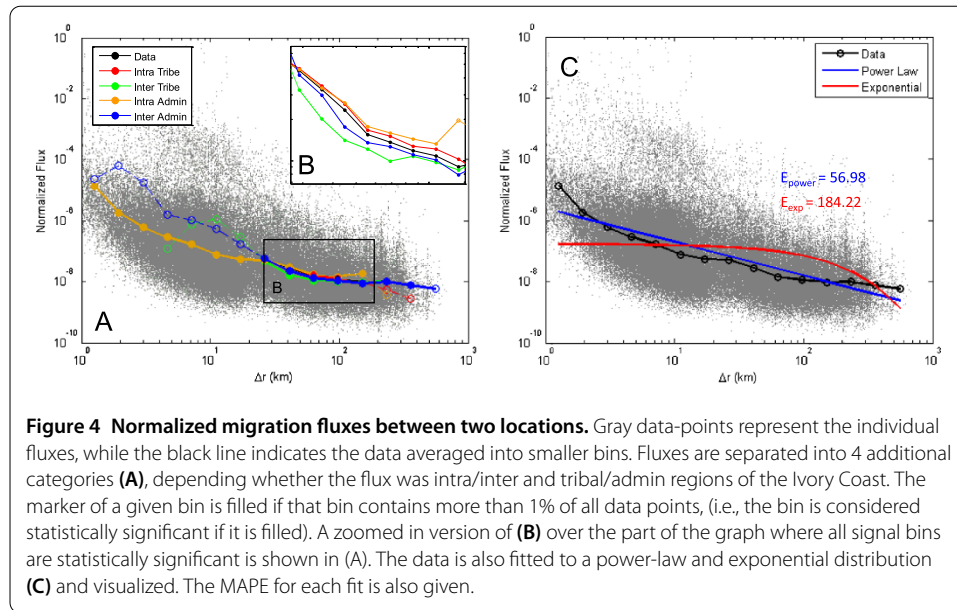
Figure 3 Hierarchical boundary and human migration visualization in Ivory Coast. (A) *Partitionings of Ivory Coast* by administrative prefectures/sub-prefectures (left) and tribal/sub-tribal communities (right). **(B)** *Intra-Inter tribal migrations*, where each node represents an individual sub-tribal community, and each link is logarithmically colored to represent the number of migrations (extracted from call records) between the two nodes.

borders. Figure 4B provides a zoomed in version of the statistically significant data points of Figure 4A and shows that for any given distance, the flux between intra-Administrative and intra-Tribal regions is consistently higher than the flux between inter-Administrative and inter-Tribal regions. One can then conclude that the fluxes crossing the tribal border are much weaker on average compared to fluxes at the same geographical distance but crossing an administrative border instead.

We also quantified the strength of these partitioning ties by computing the network modularity of the sub-tribal communities versus that of the administrative boundaries. The network modularity of the sub-tribal communities was 0.6548, in comparison to a network modularity of 0.6158 for the administrative boundaries. Given that the number of regions for both partitions (sub-tribal and sub-prefecture) is very similar, this increase of network modularity corresponding to sub-tribal communities by the definition of network modularity shows again that mobility patterns have a stronger connection in a sub-tribal country partitioning compared to that of an administrative partitioning.

Modeling human mobility

Accurately modeling human interactions between regions can present many challenges; however, effectively doing so can provide a crucial piece of information to efficiently dis-

Figure 4 Normalized migration fluxes between two locations. Gray data-points represent the individual fluxes, while the black line indicates the data averaged into smaller bins. Fluxes are separated into 4 additional categories **(A)**, depending whether the flux was intra/inter and tribal/admin regions of the Ivory Coast. The marker of a given bin is filled if that bin contains more than 1% of all data points, (i.e., the bin is considered statistically significant if it is filled). A zoomed in version of **(B)** over the part of the graph where all signal bins are statistically significant is shown in (A). The data is also fitted to a power-law and exponential distribution **(C)** and visualized. The MAPE for each fit is also given.

tribute resources, health services, etc. throughout a given area. The Gravity Model, whose origins trace back to Ravenstein's laws of migration [40], was formulated on Newton's Law of gravity, and predicts flux between a source and destination based on the populations of the source and destination, and the distance between the source and destination. More specifically, according to the gravity model, the average flux migrations from regions i to j is:

$$T_{ij} = A \frac{p_i p_j}{r_{ij}^{\gamma}}, \tag{2}$$

where i and j are origin and destination locations with populations p_i and p_j respectively at a distance of r_{ij} from each other. A is a normalization factor and γ is an adjustable parameter chosen to fit the data.

An alternative version of the gravity model using an exponential fit can be also described as:

$$T_{ij} = A p_i p_j e^{-\gamma r_{i,j}}. \tag{3}$$

To determine which model would be more appropriate we plotted the normalized flux between two locations separated by a given distance in Figure 4C. We fit these normalized migrational fluxes to both a power-law and exponential fit (blue and red lines) and obverse that the power-law distribution is much more consistent for modeling our data (which is also confirmed by comparing mean absolute percent error values reported on the plot). Therefore, for the remainder of this study we used the gravity model in Equation (2).

In order to apply the model one must inherently define the appropriate spatial resolution (i.e., the areas to model migrations between/within). By monitoring the resulting accuracy of the model, it is possible to gain insight on what type of partitioning of an area will most effectively allow for human mobility to be modeled. We started by investigating if the use of these sub-tribal communities would provide an advantage in modeling the mobility network of the Ivory Coast compared to an administrative (subprefecture) partitioning.

We computed the Gravity model using dataset *D1*, and specifically modeled the migrations for both administrative and sub-tribal partitioning of the country, and subsequently tested the accuracy of the model from the mean average percent error (MAPE) with respect to the true network of human mobility. MAPE has been shown to be a very effective measure of error in model predictions, especially when considering population forecasting [39, 41].

We compute MAPE according to:

$$M = \frac{100}{n} \sum_{t=1}^{n} \left| \frac{A_t - F_t}{A_t} \right|, \tag{4}$$

where A_t is the actual value, F_t is the forecasted value, and n is the number of data points. Therefore, an inaccurate model will subsequently yield a high MAPE value; whereas an accurate model will yield a much smaller MAPE value.
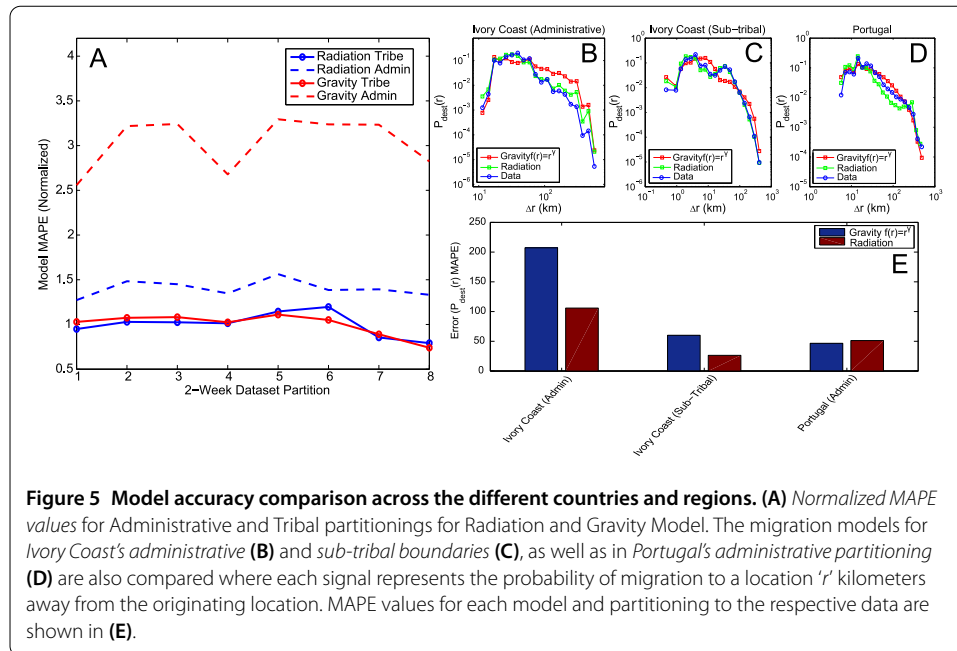
To further explore the relationship between tribal and administrative boundaries we applied an alternative approach for modeling human mobility and interaction, the Radiation Model. The Radiation Model [5] was recently proposed as a parameter free mobility model in which individuals move and interact based on the population density of the source and destination regions, and that of the surrounding regions. Using the Radiation Model, the average flux between two regions *i* and *j* is:

$$T_{ij} = T_i \frac{m_i n_j}{(m_i + s_{i,j})(m_i + n_j + s_{i,j})}, \tag{5}$$

where *i* and *j* are origin and destination locations with populations m_i and n_j respectively, at distance $r_{i,j}$ from each other, with $s_{i,j}$ representing the total population in the circle of radius $r_{i,j}$ centered at *i* (excluding the source and destination population). T_i signifies the total outgoing flux that originates from region *i*. Since, our system represents a finite space (the regions and boundaries within a country), we subsequently normalize T_i by a factor of $(1 - \frac{m_i}{M})$ where M is the total sample population in the system [42].

Figure 5A shows the normalized MAPE comparison for the sub-tribal and administrative boundaries, and demonstrates that the MAPE for Gravity Model predictions made via administrative boundaries ranged from 150% to 230% higher than that of sub-tribal communities, while the Radiation Model also yielded 20% to 50% higher MAPE values for administrative boundaries. This indicates a higher accuracy (lower MAPE) produced when sub-tribal communities were used, as opposed to administrative boundaries. This subsequently suggests that, in terms of mobility patterns of Ivory Coast, it is more effective to model mobility on a partition that accounts for tribal, cultural, and lingual differences in groups of people, as opposed to the current administratively defined country partition.

We also partitioned the mobility model predictions according to intra-tribal and inter-tribal flux in order to quantify the *strength of the connectivity of the tribes*. Quantitatively, the MAPE for the inter-tribal mobility was 11.3% higher than the MAPE of the intra-tribal mobility predictions, and supports the dominant pattern of intra-tribal migrations over inter-tribal migration which may require special consideration in terms of mobility modeling. Again, the fact that the Radiation model produces more accurate results for migrations within a single tribe compared to those between tribes suggests that the tribes themselves are playing a key role in the overall improved accuracy of the model.

Figure 5 Model accuracy comparison across the different countries and regions. (A) *Normalized MAPE values* for Administrative and Tribal partitionings for Radiation and Gravity Model. The migration models for *Ivory Coast's administrative* **(B)** and *sub-tribal boundaries* **(C)**, as well as in *Portugal's administrative partitioning* **(D)** are also compared where each signal represents the probability of migration to a location '*r*' kilometers away from the originating location. MAPE values for each model and partitioning to the respective data are shown in **(E)**.

Figures 5B-D compare the probability of migration predicted by both the Gravity and Radiation Models, to the actual migration percentages as computed from CDR data. While the CDR data is not the ground truth for migration, for example because it is a sample of the total population and is tracked only to the antenna level, it is also the basis for both models and thus represents a valid comparison for this study. As a more direct comparison of accuracy, Figure 5E provides the error (MAPE) for the models plotted in Figures 5B-D. For the Ivory Coast, these figures show the higher accuracy (i.e., lower error) of Radiation Model for both administrative and sub-tribal communities, and it shows that using the Radiation Model with sub-tribal communities provides the highest accuracy (i.e., lowest MAPE).

Figure 5D illustrates the Portuguese administrative municipality boundaries perform well for both the Radiation and Gravity Models. This may be indicative of municipal boundaries that were designed to align with cultural and social communities or that cultural and social communities have adapted to fit administrative boundaries.

However, we believe a more likely explanation is the growing homogeneity of language and culture that comes with maturing industrialization and urbanization. This is reflected in the predominance of Portuguese as the national language in Portugal, compared to the more than 60 local languages spoken in Ivory Coast.

There are several important implications from these findings.

1 Models of mobility, migration, and interaction that are conceptualized in mature and industrialized regions may not directly map to developing regions with more pronounced cultural and linguistic differences. Such models need to better account for these differences.

2 If administrative boundaries are drawn and services are placed based on models that do not accurately reflect these influences, results could include inefficiencies, leading to inequality of services (e.g., longer or less accessible commutes), and potentially discrimination and alienation of segments of the population.

3 Techniques for assessing the strength of borders in a given regional partitioning
   scheme are critical to ensuring the accuracy of mobility and migration modeling, and,
   perhaps more importantly, to enabling sound decision-making by authorities tasked
   with setting effective administrative boundaries.

In the following section, we propose novel techniques for quantifying the impact of a
regional partition scheme on model accuracy, and for assessing the strength of borders in
a given partitioning scheme.

Assessing regional affinities and border strength

In previous sections, we demonstrated the issues arising from using mobility models, such
as the Gravity and Radiation models, on regions where the partitioning scheme, such as
Ivory Coast administrative boundaries, does not reflect the regional affinities and border
strengths these models assume. We illustrated techniques to create more appropriate par-
titioning schemes, such as the tribal communities, and demonstrated the ability to achieve
higher accuracy mobility modeling using this improved partitioning.

However, it may not always be possible to simply re-draw borders. Instead, tools are
needed to assess the efficacy of an existing partitioning scheme, in terms of the affinities
within the identified borders and the strength of the borders.

In this section, we propose two novel techniques to address the above challenge. Firstly,
we present a metric to determine whether affinities exist within borders that may impact
the accuracy of mobility modeling. We test our metric on the tribal and administrative
boundaries used in the previous sections. Secondly, we propose a technique to assess the
strength of existing borders, and we demonstrate our technique on the existing adminis-
trative borders of Portugal and Ivory Coast. Our techniques use the same mobility data
used in previous sections and can be performed on any regional partitioning scheme, and
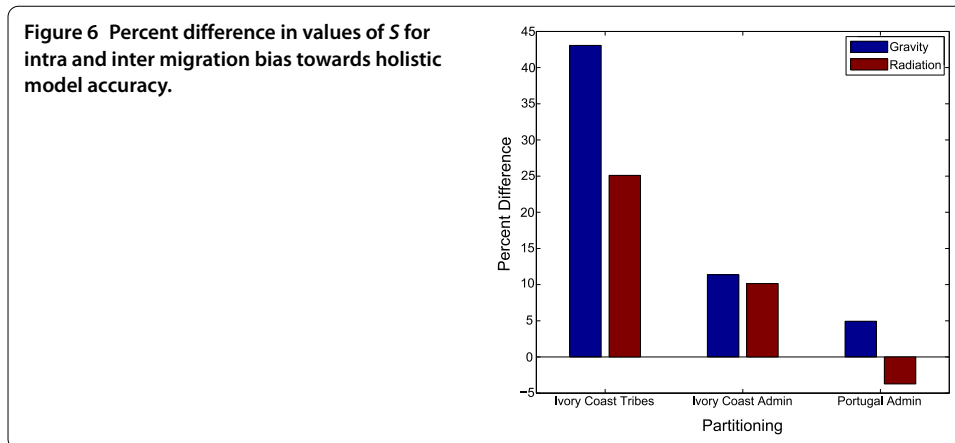thus provide valuable tools to mobility researchers and to urban planners.

Regional affinity

The accuracy of both the Gravity Model and the Radiation Model depends on the ability to
accurately model, for a given time epoch, movement from any region to any other region,
and lack of movement to another region. Inter-region movements are driven by oppor-
tunities and resources, which are reflected in the population of region, and constrained
by distance. Intra-region affinities, such as physical proximity to home, work, family, and
friends, tend to limit movement from a given region. Improperly partitioning a region to
account these affinities results in over or under predicting flux, and therefore, poor model
performance.

We propose a metric to assess whether such affinities exist, and test that metric on the
tribal and administrative boundaries used in previous sections. To compute this metric, we
segregated all migrations into two categories: intra- and inter-regional migrations. Refer-
ring to Figure 3A, an inter-region migration is a migration that crosses a color boundary.
Likewise, an intra-region migration is a migration that does not cross a color boundary.
When a model over-predicts the number of inter-region migrations, it is under-estimating
the strength of the affinities within that region.

We use S to denote the bias of a regional partitioning to over or under estimate flux
across regional boundaries. We compute S as:

$$S = \frac{\sum_1^n \frac{R}{T}}{n},$$
(6)

Figure 6 Percent difference in values of *S* for intra and inter migration bias towards holistic model accuracy.

where R is the modeled flux, T is the true flux, and n is the total number of predictions made. In other words, S represents the average ratio of modeled to predicted fluxes in the system. Therefore, relatively high values of S indicate the model is generally *over* predicting, while lower values indicate a general *under* prediction by the model. We compute separate S_{inter} and S_{intra} values, where the S_{inter} value includes only inter-region flux and S_{intra} includes only intra-region flux.

The percent difference (D) between S_{intra} and S_{inter} for all partitions provides a measure of bias. That is, the larger the value of D, the stronger the bias for over estimating inter-region flux. Figure 6 illustrates that both the Gravity and Radiation Models exhibited the most significant over estimation of inter region flux for Ivory Coast Tribes. Recall that significant over estimation of inter region flux is a result of significant under estimation of intra region affinities. Consistent with the results of previous sections, the administrative boundaries of both Portugal and Ivory Coast demonstrate substantially less bias. Thus our metric performs well in highlighting the strong affinities (in this case tribal), which must be accounted for in the model.

Border strength
Background

Just as human mobility may be constrained by affinities, it is similarly impacted by the strength of surrounding borders. Mobility models must accurately and succinctly reflect borders that may be physical, such as gates or other guards, or abstract, such as lack of opportunities. However, neither the Gravity Model nor the Radiation Model provide a means to assess border strength.

In this section, we propose a novel metric for assessing the strength of borders within a region. We demonstrate our metric by computing the strengths of the administrative borders of Portugal and Ivory Coast used in previous sections. As predicted by the Gravity and Radiation Models, we show that borders surrounding the heavily populated region of Abidjan are more penetrable than borders elsewhere in the Ivory Coast. Similarly, our results show that the borders throughout the country of Portugal are much more uniform, even showing consistency with the more heavily populated Lisbon. Our metric allows us to show that while these two cities, located in two very different regions, share a higher penetrability of borders than surrounding regions, they also exhibit significant differences in the distribution of border strength.

Methodology

To compute border strength, we start by defining the connectedness of each node (in our study, cell-tower) to each partition. The connectedness of node i to partition P is denoted by $C_{i,P}$, and computed as:

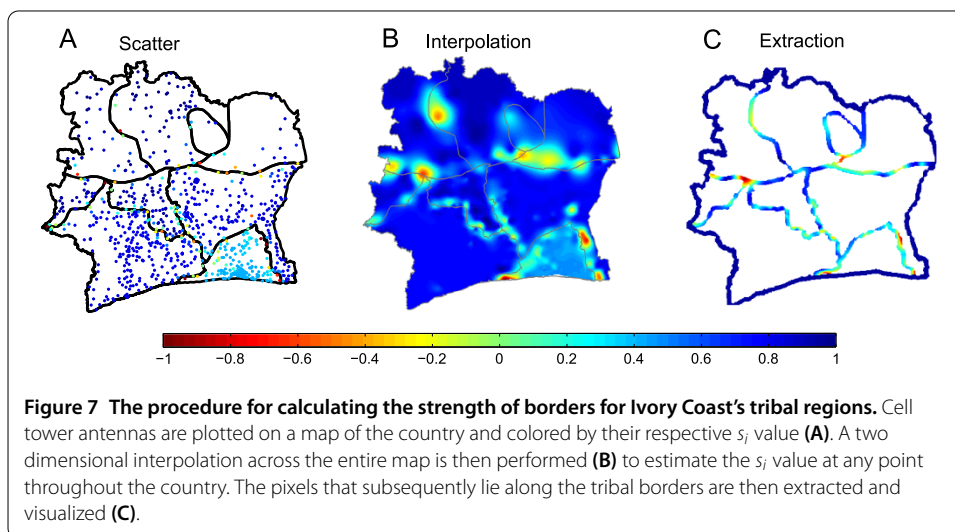$$C_{i,P} = \frac{\sum_{\substack{j \in P \\ j \neq i}} (e_{i,j} + e_{j,i})}{S_i + T_i - 2m_{i,i}}, \qquad (7)$$

where i and j are nodes, $w_{i,j}$ is the weighted directed matrix of human mobility, $m_{i,j} = \frac{w_{i,j}}{\sum w_{i,j}}$, $S_i = \sum_j m_{i,j}$, and $T_j = \sum_i m_{i,j}$. Finally, $e_{i,j} = m_{i,j} - S_i T_j$ is the difference between the actual number of migrations and the expected number of migrations from i to j.

For any given node i (located within partition P_i), if there exists a partition P_q ($i \neq q$) for which $C_{i,P_q} > C_{i,P_i}$ (i.e., node i is more strongly connected to partition P_q than to the partition in which it is located), then we say node i is not stable within partition P_i. We compute the stability of node i as s_i, where:

$$s_i = C_{i,P_i} - \max_{D \neq P_i} \{C_{i,D}\}. \qquad (8)$$

Therefore, for any given node i, $-1 \leq s_i \leq 1$. Further, $s_i = 0$ indicates node i is as strongly connected to at least one partition other than the partition P_i to which it is assigned, and thus we say node i is not stable within the partition P_i it is located. Negative values of s_i indicate node i is more strongly connected to partitions other than the partition P_i to which it is assigned (i.e., the considered partitioning is not optimal in terms of the modularity score), and positive values of s_i indicate node i is more strongly connected to its assigned partition P_i.

Figure 7A illustrates the stability values (s_i) for each node (i.e., cell tower) in Ivory Coast, relative to the partitioning imposed by administrative borders, and colored on a logarithmic scale. By performing a linear interpolation of the s_i values across the entire map of Ivory Coast (Figure 7B), we obtain an estimate of the stability across the entire region. This linear interpolation assigns a stability measure to each pixel in the XY plane ($s_{x,y}$),

Figure 7 The procedure for calculating the strength of borders for Ivory Coast's tribal regions. Cell tower antennas are plotted on a map of the country and colored by their respective s_i value **(A)**. A two dimensional interpolation across the entire map is then performed **(B)** to estimate the s_i value at any point throughout the country. The pixels that subsequently lie along the tribal borders are then extracted and visualized **(C)**.
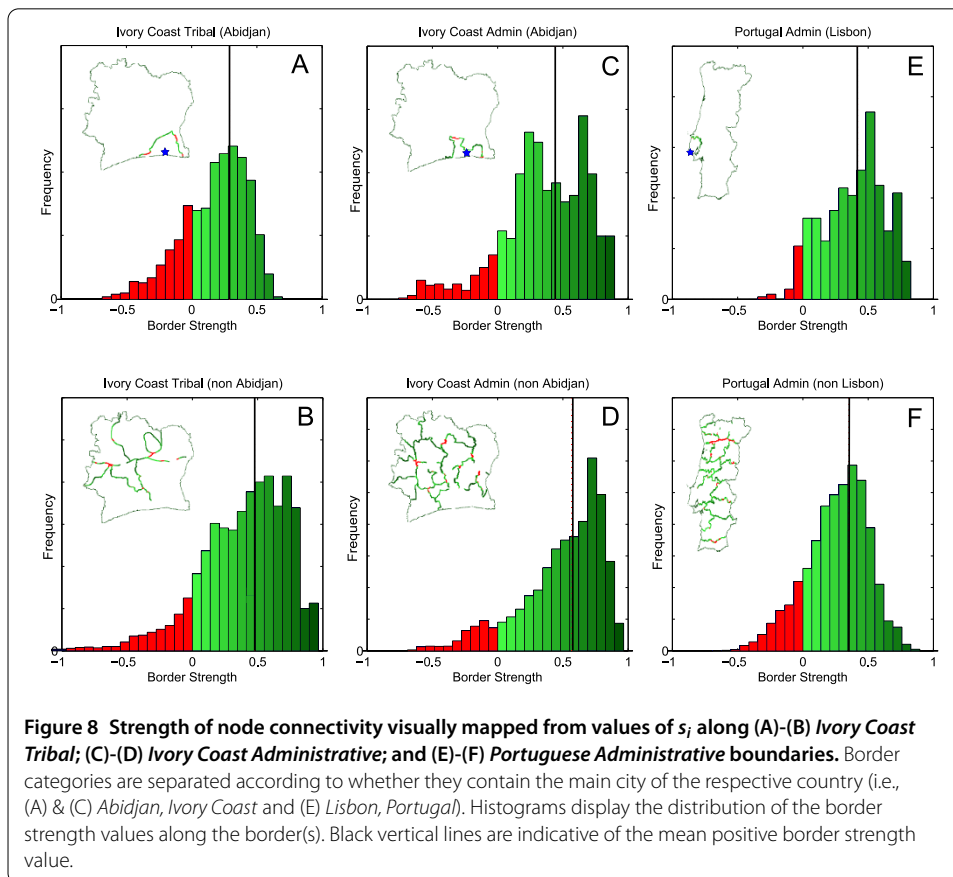
and thus enables us to compute a border strength metric for a given border as the mean of the stability values of the pixels that directly lie on that border (Figure 7C). We denote the border strength metric of any given border b as s_b. Thus $s_b = \text{mean}(s_{x_b, y_b})$ where (x_b, y_b) are all coordinates of pixels that lay on a border.

Note also that by computing s_b from the linearly interpolated values across the entire region, our approach enables a border strength metric that reflects the aggregate impact of cells that are not well connected, even if those cells are not physically close to borders.

Analysis

In addition to providing a single border strength metric by which any border may be assessed, our method enables analysis of the distribution of stability values ($s_{x,y}$) for each border. Figures 8A-D illustrate how significantly more penetrable borders surrounding the densely populated Abidjan are in comparison to other portions of the country. This higher penetrability of borders surrounding Abidjan holds for both tribal (Figure 8A) in comparison to the remainder of tribal borders (Figure 8B). This effect is again prominent with the administrative border surrounding Abidjan (Figure 8C) relative to the rest of the country (Figure 8D). Furthermore, Figure 8A illustrates the mean positive penetrability ($\overline{s_b}$) of 0.2847 for the borders surrounding Abidjan, versus 0.4767 for borders throughout the rest of Ivory Coast (Figure 8B). Also note that penetrability in Abidjan is markedly higher than in the remainder of Ivory Coast, whereas the Lisbon borders (Figures 8E-F) are only marginally less penetrable than in the remainder of Portugal. This supports our ear-

Figure 8 Strength of node connectivity visually mapped from values of s_i along (A)-(B) *Ivory Coast Tribal*; (C)-(D) *Ivory Coast Administrative*; and (E)-(F) *Portuguese Administrative* boundaries. Border categories are separated according to whether they contain the main city of the respective country (i.e., (A) & (C) *Abidjan, Ivory Coast* and (E) *Lisbon, Portugal*). Histograms display the distribution of the border strength values along the border(s). Black vertical lines are indicative of the mean positive border strength value.

lier finding that mobility in Portugal is much more uniform throughout the entire country compared to the Ivorian Coast context.

The distributions illustrated in Figure 8 also provide interesting insights into mobility in Ivory Coast and Portugal. More specifically, notice the tight clustering of border strengths for Abidjan under the Tribal partitioning (Figure 8A) versus more widely distributed border strength values for the rest of Ivory Coast. This supports our early findings that Tribal boundaries play a key role in mobility throughout the country, except in highly industrialized regions such as Abidjan.

Conclusion

Africa is a continent that has been shaped by human migration over tens of thousands of years. Indeed, migrations within and beyond the African borders have recently been shown as influencing all civilizations as we know them. However, until recently, there has been a dearth of data on the forms and patterns of migration within the nations of Africa. Moreover, much of the mobility research is based on theories that have emerged from highly industrialized nations and lack validation in the context of developing environments.

Our study has demonstrated that many of these conceptions are not necessarily applicable in the African context. We have made these differences clear by comparing our findings in Ivory Coast to one such industrialized nation, Portugal. For example, we have shown that the probability of displacement during normal commuting hours in Portugal is often nearly double that of Ivory Coast for the same time of day. Similarly, average distances traveled by commuters in Portugal is nearly double that of commuters in Ivory Coast.

While differences in the likelihood of travel and average distance travel can be attributed to quantitative differences in infrastructural support for mobility this already strongly affects the whole mobility picture leading to a number of quantitative dissimilarities. Our study shows evidence of more fundamental differences in infrastructural support for mobility, such as tribal, cultural, and lingual differences. In addition, we demonstrate that the similarity between administrative boundaries and communities detected in mobile phone data is markedly lower in Ivory Coast than in Portugal.

By identifying the tribal influence on mobility in the Ivory Coast, we were able to illuminate further differences in mobility patterns. For example, we were able to show intra-tribal migrations were much more frequent than that of inter-tribal migrations over the same distance and therefore are under or overestimated by the models. Taking this into account by exploiting our tribally aligned communities for the mobility models drastically improves modeling of human mobility in Ivory Coast. We validated this higher accuracy by computing the MAPE across all data points for both models, and found a 20% to 50% higher error for the models using administrative boundaries. We also validated our results by computing the distribution of migrations by distance migrated and found that by using this sub-tribal method of spatial units definition in modeling human mobility we were able to improve the accuracy of the models so drastically, that the Ivory Coast performed even better than its developed country-counterpart, Portugal.

We propose novel techniques for assessing the strength of borders within a regional partitioning scheme, and for assessing the impact of inappropriate partitioning on model accuracy. Our results offer improved insights on why models developed for mature and stable regions may not translate well to developing regions, and provide tools for urban planners and data scientists to address these deficiencies.

We believe that the findings of this study demonstrate important differences that exist between developing and industrialized regions. Using these two countries as an example, we are motivated to further explore these differences by considering more countries and areas with diverse cultural, economical and social backgrounds.

References
 1. Robertson C, Sawford K, Daniel SL, Nelson TA, Stephen C (2010) Mobile phone-based infectious disease surveillance system, Sri Lanka. Emerg Infect Dis 16(10):1524
 2. Gonzalez MC, Hidalgo CA, Barabasi A-L (2008) Understanding individual human mobility patterns. Nature 453(7196):779-782
 3. Ratti C, Williams S, Frenchman D, Pulselli R (2006) Mobile landscapes: using location data from cell phones for urban analysis. Environ Plan B, Plan Des 33(5):727
 4. Reades J, Calabrese F, Ratti C (2009) Eigenplaces: analysing cities using the space-time structure of the mobile phone network. Environ Plan B, Plan Des 36(5):824-836
 5. Simini F, González MC, Maritan A, Barabási A-L (2012) A universal model for mobility and migration patterns. Nature 484(7392):96-100
 6. Kang C, Sobolevsky S, Liu Y, Ratti C (2013) Exploring human movements in singapore: a comparative analysis based on mobile phone and taxicab usages. In: Proceedings of the 2nd ACM SIGKDD international workshop on urban computing. ACM, New York, p 1
 7. Blondel VD, Esch M, Chan C, Clerot F, Deville P, Huens E, Morlot F, Smoreda Z, Ziemlicki C (2012) Data for development: the d4d challenge on mobile phone data. Preprint arXiv:1210.0137
 8. Eagle N (2008) Behavioral inference across cultures: using telephones as a cultural lens. IEEE Intell Syst 23(4):62-64
 9. Eagle N, de Montjoye Y, Bettencourt LM (2009) Community computing: comparisons between rural and urban societies using mobile phone data. In: International conference on computational science and engineering, vol 4, 2009. CSE'09, pp 144-150, IEEE
10. Ahas R, Aasa A, Mark Ü, Pae T, Kull A (2007) Seasonal tourism spaces in Estonia: case study with mobile positioning data. Tour Manag 28(3):898-910
11. Berlingerio M, Calabrese F, Di Lorenzo G, Nair R, Pinelli F, Sbodio ML (2013) Allaboard: a system for exploring urban mobility and optimizing public transport using cellphone data. In: Proceedings of the third international conference on the analysis of mobile phone datasets (NetMob)
12. Liu F, Janssens D, Wets G, Cools M (2013) Profiling workers' activity-travel behavior based on mobile phone data. In: Proceedings of the third international conference on the analysis of mobile phone datasets (NetMob)
13. Zilske M, Nagel K (2013) Building a minimal traffic model from mobile phone data. In: Proceedings of the third international conference on the analysis of mobile phone datasets (NetMob)
14. Naboulsi D, Fiore M, Stanica R et al(2013) Human mobility flows in the city of Abidjan. In: 3rd international conference on the analysis of mobile phone datasets
15. Yan X-Y, Zhao C, Wang W (2013) Predicting human mobility patterns in cities. In: 3rd international conference on the analysis of mobile phone datasets
16. AfriPop: Cote D'Ivoire. http://www.clas.ufl.edu/users/atatem/index_files/CIV.htm
17. Portal do Instituto Nacional de Estatística. http://www.ine.pt
18. Voronoï G (1908) Nouvelles applications des paramètres continus à la théorie des formes quadratiques. Deuxième mémoire. Recherches sur les parallélloèdres primitifs. J Reine Angew Math 134:198-287
19. Ratti C, Sobolevsky S, Calabrese F, Andris C, Reades J, Martino M, Claxton R, Strogatz SH (2010) Redrawing the map of Great Britain from a network of human interactions. PLoS ONE 5(12):14248
20. Expert P, Evans TS, Blondel VD, Lambiotte R (2011) Uncovering space-independent communities in spatial networks. Proc Natl Acad Sci USA 108(19):7663-7668
21. Calabrese F, Dahlem D, Gerber A, Paul D, Chen X, Rowland J, Rath C, Ratti C (2011) The connected states of America: quantifying social radii of influence. In: 2011 IEEE third international conference on privacy, security, risk and trust (passat) and 2011 IEEE third international conference on social computing (socialcom), pp 223-230. IEEE
22. Sobolevsky S, Szell M, Campari R, Couronné T, Smoreda Z, Ratti C (2013) Delineating geographical regions with networks of human interactions in an extensive set of countries. PLoS ONE 8(12):81707

23. Sobolevsky S (2013) Digitale ansätze für eine regionale abgrenzung. In: Offenhuber D, Ratti C (eds) Die stadt entschlüsseln: wie echtzeitdaten den urbanismus verändern, Bauwelt fundamente, vol 150. Birkhäuser, Basel
24. Newman ME (2006) Modularity and community structure in networks. Proc Natl Acad Sci USA 103(23):8577-8582
25. Blondel VD, Guillaume J-L, Lambiotte R, Lefebvre E (2008) Fast unfolding of communities in large networks. J Stat Mech Theory Exp 2008(10):10008
26. Le Martelot E, Hankin C (2011) Multi-scale community detection using stability as optimisation criterion in a greedy algorithm. In: KDIR, pp 216-225
27. Newman ME (2006) Finding community structure in networks using the eigenvectors of matrices. Phys Rev E 74(3):036104
28. Lancichinetti A, Fortunato S (2009) Community detection algorithms: a comparative analysis. Phys Rev E 80(5):056117
29. Sobolevsky S, Campari R, Belyi A, Ratti C (2013) A general optimization technique for high quality community detection in complex networks. Preprint arXiv:1308.3508
30. Wallace DL (1983) Comment. J Am Stat Assoc 78(383):569-576
31. Rand WM (1971) Objective criteria for the evaluation of clustering methods. J Am Stat Assoc 66(336):846-850
32. Jaccard P (1901) Étude comparative de la distribution florale dans une portion des alpes et des jura. Bull Soc Vaud Sci Nat 37:547-579
33. Fowlkes EB, Mallows CL (1983) A method for comparing two hierarchical clusterings. J Am Stat Assoc 78(383):553-569
34. Heckerman D, Meila M (1998) An experimental comparison of several clustering and initialization methods. Technical report, Citeseer
35. Hubert L, Arabie P (1985) Comparing partitions. J Classif 2(1):193-218
36. Larsen B, Aone C (1999) Fast and effective text mining using linear-time document clustering. In: Proceedings of the fifth ACM SIGKDD international conference on knowledge discovery and data mining. ACM, New York, pp 16-22
37. Onnela J-P, Arbesman S, González MC, Barabási A-L, Christakis NA (2011) Geographic constraints on social network groups. PLoS ONE 6(4):16939
38. Cote d'Ivoire - Encyclopedia Britannica Academic Edition Inc. http://www.britannica.com/EBchecked/topic/139651/Cote-dIvoire
39. Swanson DA, Tayman J, Bryan T (2011) Mape-r: a rescaled measure of accuracy for cross-sectional subnational population forecasts. J Popul Res 28(2-3):225-243
40. Ravenstein EG (1885) The laws of migration. J Stat Soc 48(2):167-235
41. Tayman J, Swanson DA (1999) On the validity of mape as a measure of population forecast accuracy. Popul Res Policy Rev 18(4):299-322
42. Masucci AP, Serras J, Johansson A, Batty M (2013) Gravity versus radiation models: on the importance of scale and heterogeneity in commuting flows. Phys Rev E 88(2):022812