Ubiquitous Sensing for Mapping Poverty in Developing Countries

Christopher Smith
ICRI: Cities
London, UK
chris.smith@ucl.ac.uk

Afra Mashhadi
Bell Labs, Alcatel-Lucent
Dublin, Rep. of Ireland
afra.mashadi@alcatel-lucent.com

Licia Capra
University College London
London, UK
l.capra@ucl.ac.uk

ABSTRACT
Social surveys and censuses offer a good indication of poverty and inequality in a country. However, due to the expenses associated with data collection, the granularity and frequency of such information is often limited. In developing countries in particular, census data may be collected very infrequently, thus failing to accurately reflect the changes associated with a growing economy. In this paper, we propose to use ubiquitous sensing as a proxy for estimating socioeconomic indicators and analyse aggregated mobile phone communication data in Côte d'Ivoire. We discover a number of features that show a strong correlation with poverty indicators. We then demonstrate how these features can be used to provide poverty estimates at a spatial resolution finer than previously available.

INTRODUCTION
Social surveys and censuses periodically collected by National Statical Institutes contain valuable information describing the social and economic well being of a country and the relative health of different areas. Such data is used by policy-makers and agencies to guide the formulation and implementation of policies and programs that aim to improve the life of the citizens. Poverty maps derived from survey data, and spatial descriptions of the distribution of poverty are most useful when they are finely disaggregated (i.e., when they represent small geographic units, such as cities, towns or villages), and more importantly, when they are most up to date. Spatially rich and temporally accurate knowledge of socioeconomic indicators would help in alleviating poverty by enabling efficient investment in infrastructure and consequently mitigating against the detrimental effects of poverty and inequality. However, this form of data collection is known to be an onerous task due to the cost involved, especially for nations where political instability and a weak economy exacerbate the problem.

Côte d'Ivoire is an example of a developing country which has suffered recent political strife and economic turmoil. Agriculture employs roughly 68% of its total population with Côte d'Ivoire being the world's largest producer and exporter of cocoa beans and a significant producer and exporter of coffee. Although rich in agriculture and natural resources (e.g., diamond), the economy is highly sensitive to fluctuations in international prices for these products. Furthermore, recent events, including civil war, have resulted in a loss of foreign investment and economic contraction. In late 2011, Côte d'Ivoire's economy began to recover from a severe downturn in the first quarter of the year that was caused by widespread post-election conflict. In June 2012 the World Bank announced $4.4 billion in debt relief for Côte d'Ivoire under the Highly Indebted Poor Countries Initiative. Côte d'Ivoire's long term challenges are known to include political instability and degrading infrastructure [1].

Given this state of affairs, it is perhaps not surprising that no data pertaining to a full survey of the country's population appears to have been made available since the late 1990s. To address this problem, we propose the use of ubiquitous sensing as an alternative to the traditional method of collecting sociodemographic data through census and social surveys. Ubiquitous sensing refers to the passive collection of peoples' digital footprints (e.g., location based social networking check-ins, phone calls, etc.) which can provide a detailed picture of human mobility and communication. If we are to provide a viable alternative, or at least a useful complement, to traditional censuses, we need to ensure that data is sourced uniformly from the population, with minimum bias. Online social network and location based service data is likely to be lacking in this regard, as they suffer known demographic biases and uptake of such services tends to be clustered geographically. In this paper, we show how Call Detail Records (CDRs) can be mined in order to derive proxies for poverty indicators, which can then be used to estimate poverty on a continuous basis and at low cost, as opposed to the slow iteration of census survey cycles. Côte d'Ivoire contains more than 17 million mobile phone users (around 77% of the total population) and is well developed by African standards, being ranked 51st in the world [1]. With this high penetration rate there is significant potential for methods exploiting ubiquitous sensing to have a real impact.

[1]CIA World Factbook-`https://www.cia.gov/library/publications/the-world-factbook/geos/iv.html`

Submitted to D4D @ NetMob 2013

We first describe the features we mine from the data and report their correlation with poverty as measured by the Multiple Poverty Index. We then demonstrate how these features could be used to build a regression model that estimates poverty at a finer level of granularity, potentially enabling policymakers and agencies to efficiently allocate limited resources to avert and re-balance inequality and invest in infrastructure where needed most.

RELATED WORK

There is today an unprecedented amount of location based digital footprint data, such as geo-tagged tweets, Foursquare check-ins and CDRs, which has been the subject of much research aiming to understand the dynamics of human mobility and communication on a scale not previously possible. Noulas *et al.* [17] study urban mobility patterns of people in metropolitan cities by analysing the check-ins of a large sample of Foursquare users. They find that human mobility appears to obey a universal law which isolates as a key component the rank-distance, which factors in the number of places between origin and destination, rather than pure physical distance. Cheng *et al.* [8], investigate the effect of contextual factors such as population and social status on the mobility patterns of citizens through their digital footprints obtained from Foursquare check-ins. They observe that in addition to geographic and economic factors, social status is highly coupled with one's mobility in the city. In particular, they show that people in wealthy cities travel more frequently to distant places than people in less rich cities. Kramer [14] found that the difference between the number of positive and negative words used in Facebook status updates covaries with self-reported "satisfaction with life" in the US. Similarly, Quercia *et al.* found that sentiment expressed in tweets [18] and the topic of tweets [19] in London, aggregated by the area associated with the tweet or Twitter profile, correlates with socioeconomic deprivation of that area. A limitation noted in these works, however, is the large demographic bias of online social networking services. The majority of Twitter users are male, under 35 and with a relatively high income. Mislove *et al.* [16] also suggest that the ethnicity of twitter users, in the US at least, is not representative of the general population. Similarly, although Facebook has a more even gender distribution, in the UK around 60% of users are less than 35 years of age.[2]

In order to avoid such a population bias an alternative data source that is more representative of the population is required. One such source are CDRs, which have been extensively studied for a broad range of purposes, from understanding human mobility [6, 11, 7, 26] to land use identification and urban planning [4, 24, 20]. Various ways of characterising geography based on the traffic of mobile phones and their users' trajectories have been examined. Specific to understanding the relation between CDRs and socioeconomic factors, there has been only a handful of works in the literature [25, 10, 5, 9]. Eagle *et al.* [9] measured the communication diversity from fixed line phone call records in England, and found that higher diversity (i.e., the more evenly dispersed a person's communication between people and places) correlates with socioeconomic deprivation aggregated to telephone exchange areas. Blumemstock [5] looked at the relation between users' demographics (collected through personal interviews) and their mobile phone usage from a sample of employees from a company in Rwanda. Observations include that gender and social status of the users had a direct correlation with the volume of their call activity. The closest work to that presented here is research undertaken by Soto *et al.* [25] and Frias-Martinez *et al.* [10], in which the authors have proposed models to infer and predict the socioeconomic indicators of a region. Specifically, [25] proposes a Support Vector Machine model operating on 279 features of individual users' CDR to infer the socioeconomic level of census regions. They used features categorised into 69 *behavioural* (such as total number of calls), 192 *social* (such as number of contacts) and 18 *mobility* features (such as distant travelled). The authors report the performance accuracy rate of around 80%. Finally, [10] has extended [25] to provide forecasts of socioeconomic factors. The drawback to this approach is that by including so many features and their interactions in a complex model policymakers are presented with a 'black-box' predictor, with little hope of understanding how the estimates are reached. Arguably, for such predictions to play a role in the decision making process, it is vital that it can be understood how they were formed. Furthermore, many of the features used in these works require detailed knowledge of individual behaviour, which for privacy reasons may not be readily available.

For these reasons our approach differs from the above works in two important ways: i) we consider only CDRs aggregated by the antennas through which the calls are connected; and ii) our results suggest that far less input variables are needed to infer poverty levels from the aggregated CDRs. We thus avoid the privacy issues associated with individual user data, and allow for a more detailed understanding of how our estimates are formed.

MINING CALL DATA RECORDS

In order to infer poverty levels of areas from human communication patterns, we require a communication dataset representative of the population as well as a ground-truth dataset of poverty information to validate our approach against. We describe each of these datasets next.

Call Records

We obtained a dataset of anonymised voice calls between five million of Orange customers in Côte d'Ivoire between December 1st 2011 and April 28th 2012[3]. Orange is the second main provider of mobile services in Ivory Coast, keeping 48% of market share as well as possessing the largest

[2] `http://www.insidefacebook.com/2010/06/08/whos-using-facebook-around-the-world-the-demographics\-of-facebooks-top-15-country-markets/` - retrieved 29/05/2012

[3] As part of D4D challenge by Orange, see `http://www.d4d.orange.com/`

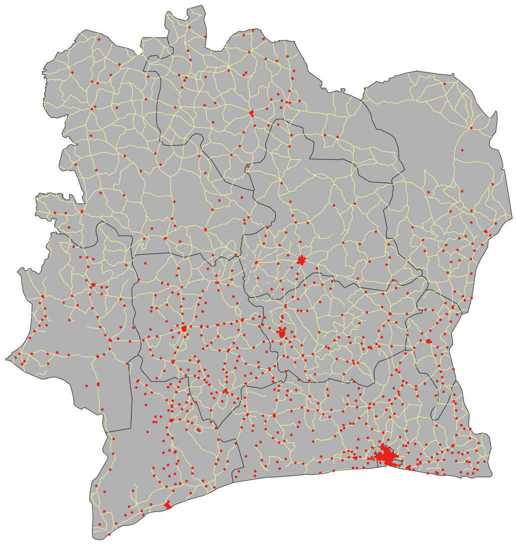

Figure 1. Spatial Distribution of Antenna in Côte d'Ivoire

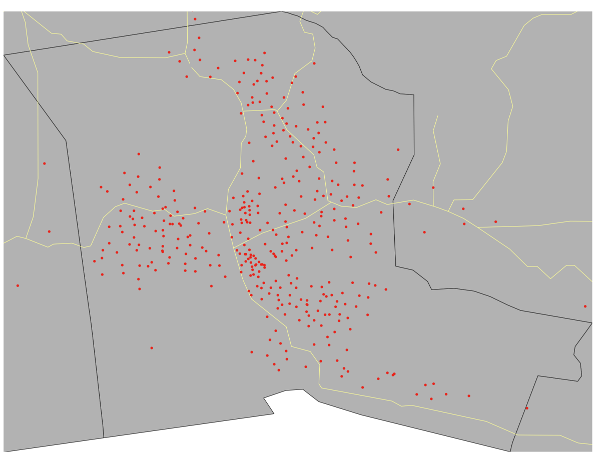

Figure 2. Spatial Distribution of Antenna in the city of Abidjan.

network of retail stores[4]. For the purpose of this study we used the subset data of antenna to antenna traffic, which contains hourly aggregated volume and duration of calls between pairs of antenna. Figure 1 shows the location of the antennas within eleven regional boundaries of Côte d'Ivoire. Notice that in the south there is a dense concentration of antennas in and around the secondary capital, Abidjan, which has a higher population density than the rest of the country and is where most economic activity and trading takes place. We first mapped the antennas to the regions in which they fall before summing volume and duration of calls made between each pair of regions. Some antennas were found to have the same coordinates, and it is unknown to the authors whether these are errors or whether there are genuinely more than one antenna in the same location. However, for our purposes what is important is the location of source and target, therefore we reassigned the identifiers of antennas with duplicated locations.

Multidimensional Poverty Index

In order to study the correlation between the call detail records and poverty, we also require a ground-truth dataset of poverty levels for areas in Côte d'Ivoire. For this purpose, we used the Multidimensional Poverty Index created by the University of Oxford[5]. The index incorporates a range of indicators in order to reflect the complexity of poverty and better inform policies aimed at relieving it. In addition to the single valued index and population estimates, the MPI contains several indicators that aim to capture people's experience of deprivation, such as poor health, lack of education, inadequate living standard, low income, disempowerment and threat of violence. Although detailed in terms of its coverage of the various facets of poverty, the Côte d'Ivoire MPI is derived from survey data from the year 2005. The temporal lag between our ground truth data and the mobile phone data introduces a limitation to our findings and the methods we present. This will be discussed further in the discussion section. Figure 3 depicts the aggregated MPI for eleven regions of Côte d'Ivoire, where the darker colour indicates higher MPI thus higher poverty.

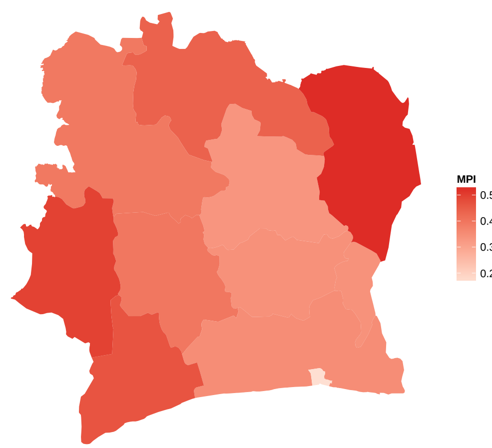

Figure 3. Map of 11 regions showing Multidimensional Poverty Index. Lighter colour indicates higher poverty.

Testing Hypotheses

We next formulate a number of hypotheses and derive features of the flow data to test them.

[4] `http://www.orange.com/en/group/global-footprint/countries/Group-s-activities-in-Ivory-Coast`

[5] `http://www.ophi.org.uk/policy/multidimensional-poverty-index/`

Activity
The first set of features are simple aggregates of the flows between regions. We expect to find that the level of mobile communication activity within a region will reflect its social and economic activity, and thus its level of prosperity [1]. We find strong negative correlations between the total outgoing volume ($r = -.774$, p–value= $.005$) and duration ($r = .791$, p–value= $.004$) of calls within a region and its MPI score, thus confirming that aggregated communication activity provides a simple proxy for poverty level. These aggregates highlight the relationship between communication activity and poverty in a region, however, we also aim to capture the relationship between poverty and the interactions between regions. We therefore investigate further hypotheses involving more complex features. Furthermore, it may not always be the case that these simple aggregates will provide accurate enough estimates of poverty, particularly at finer levels of granularity.

Gravity Residuals
The next set of features involves using a gravity model to estimate the flow between the centroids of the eleven regions. First introduced by Zipf in 1946 [28], gravity models rest on the hypothesis that the size of flow between two areas is proportional to the mass (i.e., population) of those areas, but decays as the distance between them grows. Despite some criticisms (see for example [22]), the model has been successfully used to describe macro scale interactions (e.g., between cities, and across states), using both road and airline networks [3, 12] and its use has extended to other domains, such as the spreading of infectious diseases [2, 27], cargo ship movements [13], and to model intercity phone calls [15]. We hypothesise that the difference between observed and expected flows between areas reflects the level of social and economic activity in those areas, and thus will be related to poverty.

We use the following equation to find the expected flows between regions,

$$F_{u,v}^{est} = g\frac{m_u m_v}{d_{u,v}^2} \tag{1}$$

where m_u is the population of region u taken from 2010 estimates and $d_{u,v}$ is the as-the-crow-flies distance between centroids of regions u and v. We then take two error measures at the areal level: firstly, we find the pairwise residual (i.e., the error between the real and estimated flow on each link), then the *link residual* is the average of a region's links; secondly, we sum the total observed and estimated incoming/outgoing flows for each region and measure the difference, or the *sum residual*. In previous work we modelled the flow of passengers in London's rail system in a similar fashion, and found that the gravity residuals were related to deprivation of neighbourhoods. Results of all correlations are presented in Table 1. We see strong, negative correlations between a region's MPI and both its sum residual and link residual, meaning that when flows between two areas are overestimated (negative residual) the poverty level is likely to be higher, and conversely, when flows are underestimated (positive residual) the poverty level is likely to be lower.

Diversity
Our next set of features aims to capture the opportunity for development afforded by an advantageous position in an information flow network. By studying a the social network represented by a fixed line call dataset, Eagle *et al.* [9] showed that the average *diversity* of the social connections of people living in a neighbourhood correlates strongly with the level of socioeconomic deprivation (a concept closely related to poverty) in that neighbourhood. In this work we are constrained by the aggregation of the call records to antenna and are unable to look directly at the underlying individual social network. Instead, we hypothesise that the diversity of a region's connections to other regions will also reflect the level of poverty in the region. We thus take two measures of a region's diversity: first, the *degree* of the region, and the second, simply termed *diversity*, which is found using the following formula from [9],

$$diversity(i) = \frac{-\sum_j v_{i,j}\log(v_{i,j})}{\log(k_i)} \tag{2}$$

where $v_{i,j}$ is the fraction of antenna i's flow that goes to j, and k_i is the unweighted degree of i. The feature we name *degree* is the number of links connecting each region whose weights are above the 8th decile of the overall weight distribution. We test all deciles and found the 8th to give the strongest correlation with MPI. Thus, *degree* represents the number of heavily weighed connections a region has. Both degree and diversity were first calculated per antenna, then the average of all antenna within a region was taken. As with the gravity model residuals, we see a strong negative correlation with MPI, which shows that the more diverse a region's connections, the lower the poverty level is likely to be.

Introversion
Finally, we hypothesise that a region's level of *introversion* may be a signal of its poverty level. In other words, if an area has relatively fewer connections to other regions compared to the number of connections that exist within it, the less it will be able to benefit from new sources of opportunity arising further afield. In conjunction with our first hypothesis, that higher activity reflects lower poverty, for two regions with equal activity we would expect that with lower introversion to have the lower poverty level. This is similar in spirit to the theory of open economies, albeit on a different scale, which expects nations that close their borders to international trade to fair less well than those that are more open [21]. It is also related to the idea of diversity of connections, except that we now take into account space and consider only a binary relationship, that is, the ratio of self-flow to total flow. We first calculated the introversion of antennas with the following equation, and then found the average introversion of all antennas within each region.

$$introversion(i) = \frac{f_{i,i}}{\sum_{i \neq j} f_{i,j}} \tag{3}$$

where $f_{i,j}$ is the flow between antennas i and j. This measure produces values in the range $[0, 1]$ where 0 means completely extroverted and 1 means completely introverted. Again, the introversion of regions correlates strongly with MPI, this time

Hypothesis	Feature	Pearson's r	95% Confidence Interval	p-value
Activity	total volume	-.777	-.939, -.331	.005
	total duration	-.783	-.941, -.345	.004
Gravity Residuals	link volume residual	-.781	-.940, -.340	.005
	link duration residual	-.525	-.856, .109	.097
	sum volume residual	-.804	-.947, -.393	.003
	sum duration residual	-.822	-.952, -.437	.002
Diversity	diversity volume	-.834	-.956, -.469	.001
	diversity duration	-.848	-.960, -.506	.001
	degree volume	-.787	-.942, -.354	.004
	degree duration	-.750	-.931, -.274	.008
Introversion	volume introversion	.793	.368, .944	.004
	duration introversion	.795	.373, .945	.003

Table 1. Correlations between MPI of 11 regions in Côte d'Ivoire and features derived from mobile phone data.

positively, confirming our hypothesis that areas with higher levels of poverty also tend to be more introverted.

We have seen that a number of different features of communication patterns correlate strongly with MPI at the regional level. However, to be effective in targeting areas most in need of help and aiding the policymaker's decision process, we need to be able to provide estimates at a much finer level of granularity. What follows is a demonstration of the kind of estimates which could be derived from the features described above.

Estimating Poverty

In this section we use the features we derived in the previous section to estimate the level of poverty at a finer granularity. Unfortunately, due to the data collection limitation there is no poverty information available at this level with which we can validate the results, therefore we intend this exercise to be taken as a demonstration of the way in which communication data could be used. At this point it may be objected that at a finer level of granularity we would expect to see weaker correlations between poverty and our flow features. This may well be the case, however, as we would also be working with a larger number of data points which would allow us to combine several features into a more sophisticated model, one can argue that the predictive accuracy would actually increase.

To demonstrate the potential for using communication data to estimate poverty level at a finer level of granularity we use diversity of call duration as this had the strongest correlation in our previous experiments. We first derive a linear model using ordinary least squares regression,

$$MPI_u^{est} = 1.346 - 1.385 \times diversity(u) \tag{4}$$

Figure 4 depicts the MPI level for the eleven large regions of Côte d'Ivoire as predicted by this model, where the darker areas indicate higher estimated poverty level.

We can then use this model to estimate poverty levels at the sub-prefecture level, of which there are 255 in Côte d'Ivoire. Figure 5 shows the choropleth of the estimates for sub-prefectures. Notice the change in spatial pattern compared to the regional map in Figure 3. The coarser grained map depicts poverty increasing as we radiate out from the city of Abidjan. Instead, our finer grained estimates suggest that the

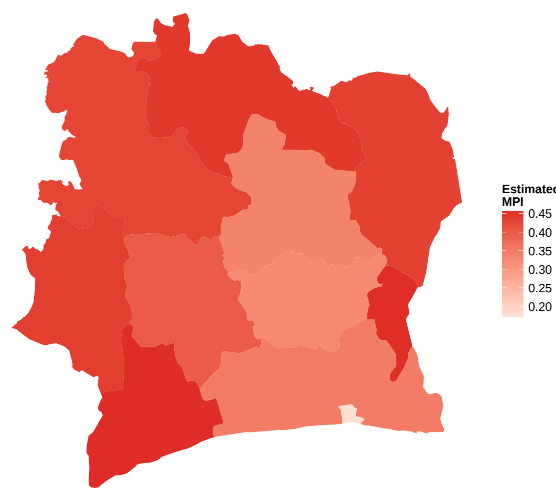

Figure 4. Poverty map estimated based on the link diversity antennas in 11 regions.

South-east of the country may contain areas of high poverty near Abidjan and conversely the North-west may contain areas of low poverty. The grey spaces in the choropleth indicate sub-prefectures for which we cannot obtain estimated poverty levels since they contain no antennas. Estimates could be extended to these regions by borrowing information from neighbouring areas, using tessellation to determine the effect of nearest antennas, or some combination of the two.

DISCUSSION

We have demonstrated the potential of CDR data to provide an invaluable source of poverty estimates, even without knowledge of individual behaviour. We have uncovered several features of communication patterns among mobile phone users in Côte d'Ivoire that track poverty of regions as defined by the Multidimensional Poverty Index. Our results have important implications for policymakers and agencies working in countries which lack the resources to manually collect so-

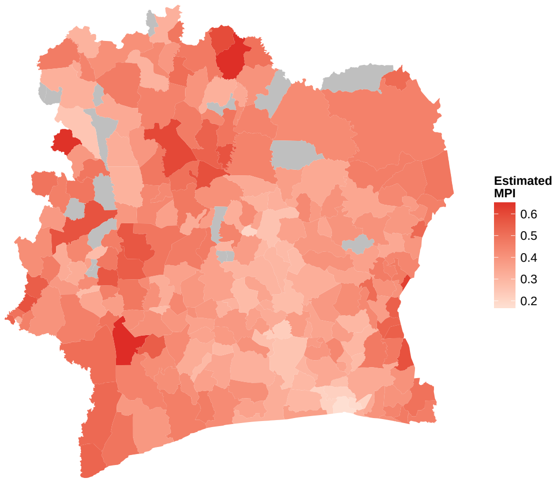

Figure 5. Poverty map in finer granularity estimated based on the diversity of connections between antenna.

cioeconomic data. Indeed, tools built upon the methods we have described would be a useful augmentation to socioeconomic data collection processes in any country. The cost of producing estimates from passively and automatically collected communication data is negligible compared to that of manual surveying, thus the main barrier to obtaining up to date poverty estimates has been removed. Côte d'Ivoire is a perfect example of a country in which timely and accurate information regarding poverty is severely lacking. In cases such as this the ability to obtain estimates of poverty levels on a continuous basis would represent a vast improvement. Limited resources could be allocated in much more efficient manner thereby helping to alleviate some of the detrimental effects of poverty and inequality.

However, we also note that the problem of a lack of up to date and spatially accurate socioeconomic data also represents a limitation to our results. In order to discover proxies for poverty indicators we require knowledge of those indicators, and to be confident that those proxies accurately track poverty level we need the proxy data and ground truth data to be close in space and time. Instead we are forced to work with a lag of 7 years between the poverty data we use as ground truth and the mobile phone data from which we derive our proxies. However, although this temporal lag will undoubtedly affect the accuracy of predictive models based on our proxies, such as the simple linear model we present above, we argue that the legitimacy of the methods we have developed is not compromised. Rather, we would only expect the accuracy and utility of the methods to increase were this lag removed.

Furthermore, it may not be strictly necessary to undertake a new, comprehensive, manual survey in order to build highly accurate predictive models. Previous work has shown that by taking a machine learning approach we may estimate socioeconomic indicators by training on a sample of census data [23]. Indeed, our own more recent experiments on deprivation in London neighbourhoods suggest that by incorporating spatial properties, the size of the required training sample could be as little as 10% of areal units.

A valuable extension to our work would then be to obtain more up to date socioeconomic data, perhaps by working with agencies on the ground to collect data from various locations. This would allow us to build a clearer picture of the relationship between communication patterns and poverty at a higher resolution. As part of our future work, we improve upon the results obtained thus far by exploring variations on the features we have defined. For example, in the gravity model rank-distance or the cost and duration of travel may be more appropriate than straight line distance between two regions, since they better reflect the considerations people make. In addition, we will explore different ways of assigning variables pertaining to antennas to the area around them. At present, antennas near areal borders are treated as if they only relate to the area in which they fall. To overcome this we could use population weighted tessellation for example. Subject to data availability, we also aim to study the generalisation of models built using the methods we have presented by comparing results in other countries, and finally, by obtaining longer term data we can investigate changes in communication patterns as changes occur in the socioeconomic well being of areas, thus helping to tease out causal relationships.

REFERENCES

1. Aker, J., and Mbiti, I. Mobile phones and economic development in africa. *Center for Global Development Working Paper*, 211 (2010).

2. Balcan, D., Colizza, V., Gonalves, B., Hu, H., Ramasco, J. J., and Vespignani, A. Multiscale mobility networks and the spatial spreading of infectious diseases. *Proceedings of the National Academy of Sciences 106*, 51 (2009), 21484–21489.

3. Barrat, A., Barthélemy, M., Pastor-Satorras, R., and Vespignani, A. The architecture of complex weighted networks. *Proceedings of the National Academy of Sciences of the United States of America 101*, 11 (Mar. 2004), 3747–52.

4. Becker, R., Caceres, R., Hanson, K., Loh, J., Urbanek, S., Varshavsky, A., and Volinsky, C. A tale of one city: Using cellular network data for urban planning. *Pervasive Computing, IEEE 10*, 4 (2011), 18–26.

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

6. Calabrese, F., Pereira, F., Di Lorenzo, G., Liu, L., and Ratti, C. The geography of taste: analyzing cell-phone mobility and social events. *Pervasive Computing* (2010), 22–37.

7. Candia, J., González, M., Wang, P., Schoenharl, T., Madey, G., and Barabási, A. Uncovering individual and collective human dynamics from mobile phone records. *Journal of Physics A: Mathematical and Theoretical 41*, 22 (2008), 224015.
8. Cheng, Z., Caverlee, J., Lee, K., and Sui, D. Exploring millions of footprints in location sharing services. In *Proc.of AAAI ICWSM* (2011).
9. Eagle, N., and Macy, M. Network Diversity and Economic Development. *Science 1029* (2010).
10. Frias-Martinez, V., Soguero-Ruiz, C., Josephidou, M., and Frias-Martinez, E. Forecasting socioeconomic trends with cell phone records. In *3rd ACM Symposium on Computing for Development* (2013).
11. Girardin, F., Calabrese, F., Fiore, F., Ratti, C., and Blat, J. Digital footprinting: Uncovering tourists with user-generated content. *Pervasive Computing, IEEE 7*, 4 (2008), 36–43.
12. Jung, W., and Wang, F. Gravity model in the Korean highway. *EPL (Europhysics Letters) 81* (2008).
13. Kaluza, P., Kölzsch, A., Gastner, M. T., and Blasius, B. The complex network of global cargo ship movements. *Journal of the Royal Society, Interface / the Royal Society 7*, 48 (July 2010), 1093–103.
14. Kramer, A. D. I. An Unobtrusive Behavioral Model of Gross National Happiness. In *Proceedings of the 28th ACM CHI* (2010), 287–290.
15. Krings, G., Calabrese, F., Ratti, C., and Blondel, V. D. Urban gravity: a model for inter-city telecommunication flows. *Journal of Statistical Mechanics: Theory and Experiment 2009*, 07 (May 2009), L07003.
16. Mislove, A., Lehmann, S., Ahn, Y., and Onnela, J. Understanding the Demographics of Twitter Users. *Fifth International AAAI* (2011), 554–557.
17. Noulas, A., Scellato, S., Lambiotte, R., Pontil, M., and Mascolo, C. A tale of many cities: universal patterns in human urban mobility. *PloS one 7*, 5 (2012), e37027.
18. Quercia, D., Ellis, J., Capra, L., and Crowcroft, J. Tracking Gross Community Happiness from Tweets. In *Proceedings of ACM CSCW 2012* (2012).
19. Quercia, D., Seaghdha, D. O., and Crowcroft, J. Talk of the City : Our Tweets , Our Community Happiness. In *Proc.of AAAI ICWSM* (2012).
20. Ratti, C., Williams, S., Frenchman, D., and Pulselli, R. Mobile landscapes: using location data from cell phones for urban analysis. *ENVIRONMENT AND PLANNING B PLANNING AND DESIGN 33*, 5 (2006), 727.
21. Sachs, J. D., and Warner, A. M. Sources of slow growth in african economies. *Journal of African Economies 6*, 3 (1997), 335–376.
22. Simini, F., González, M. C., Maritan, A., and Barabási, A.-L. A universal model for mobility and migration patterns. *Nature* (Feb. 2012), 8–12.
23. Smith, C., Quercia, D., and Capra, L. Finger On The Pulse: Identifying Deprivation Using Transit Flow Analysis. *Proceedings of ACM CSCW 2013* (2012).
24. Soto, V., and Frías-Martínez, E. Automated land use identification using cell-phone records. In *Proceedings of the 3rd ACM international workshop on MobiArch*, ACM (2011), 17–22.
25. Soto, V., Frias-Martinez, V., Virseda, J., and Frias-Martinez, E. Prediction of socioeconomic levels using cell phone records. *User Modeling, Adaption and Personalization* (2011), 377–388.
26. Toole, J., Ulm, M., González, M., and Bauer, D. Inferring land use from mobile phone activity. In *Proceedings of the ACM SIGKDD International Workshop on Urban Computing*, ACM (2012), 1–8.
27. Viboud, C., Bjø rnstad, O. N., Smith, D. L., Simonsen, L., Miller, M. A., and Grenfell, B. T. Synchrony, waves, and spatial hierarchies in the spread of influenza. *Science (New York, N.Y.) 312*, 5772 (Apr. 2006), 447–51.
28. Zipf, G. The P 1 P 2/D hypothesis: On the intercity movement of persons. *American sociological review 11*, 6 (1946), 677–686.