Bringing Visibility to Rural Users in Cote d'Ivoire

Mariya Zheleva, Paul Schmitt, Morgan Vigil and Elizabeth Belding

Department of Computer Science
University of California, Santa Barbara
{mariya, pschmitt, mvigil, ebelding}@cs.ucsb.edu

ABSTRACT
Cellular networks are often the first telecommunications infrastructure in developing regions. By studying cellular network traffic, researchers gain insight into how technologies can be used to access services critical to further development. In this work, we approach a cellular traffic dataset provided by Orange in Cote d'Ivoire with the goal of identifying distinctions between urban and rural use of cellular infrastructure. We report on a number of interesting differences between urban and rural usage of cellular infrastructure. For instance, 70% of calls that originate in rural areas occur within the vicinity of the same antenna, whereas the same is true for only 23% of calls with urban origin. We are compelled to conclude that development efforts for rural areas might be implemented differently from development efforts in urban areas based on divergent use of current cellular infrastructure.

Categories and Subject Descriptors
C.2 **[Computer-Communication Networks]:** Network Operations;

General Terms
Measurement.

Keywords
Cellular networks, rural and urban area usage, analysis, Cote d'Ivoire.

1 INTRODUCTION
Cellular networks have become one of the most prevalent means of communication worldwide. According to the International Telecommunications Union, in 2011 there were 85.7 cellular subscriptions per 100 people [11]. This number has been gradually increasing over the last ten years, with growth over the last five years predominantly driven by subscriptions in developing countries. Although GSM handsets comprise the bulk of mobile technologies in the developing world, the prevalence of smart mobile technologies is an impending reality [12].

While smart mobile technologies provide increased opportunities for access to health care [15], education [2], and economic resources [13], they also require the continued advancement of both cellular and broadband infrastructure in the developing regions where they are deployed. This makes analysis of mobile data and information flow crucial for the continued development of data infrastructures in developing countries.

ICTD 2013, Dec 07-10 2013, Cape Town, South Africa
ACM 978-1-4503-1907-2/13/12.
DOI

In this paper, we approach a cellular network dataset from Cote d'Ivoire with an emphasis on understanding how population density impacts the use of cellular infrastructure. Observations from Burrell's Invisible Users undergird our analysis [4]. Burrell uses the term "invisible users" to characterize marginalized users who are invisible due to their lack of accommodation by those in power. In the case of our analysis of cellular network use in Cote d'Ivoire, we consider invisible users to be those located in rural areas. This is due to the utilization of a commercial infrastructure that the users do not have the capacity (in terms of population density) to subsidize. This paper makes the following contributions (i) we identify stark differences in cellphone usage patterns in rural and urban areas; (ii) we find that communication patterns between antennas in similar population densities differ from those between antennas in different population densities; (iii) we recommend ways to extend infrastructure to better serve rural users in Cote d'Ivoire.

2 RELATED WORK
Previous research related to our analysis can be divided into three categories: (i) geography of mobile communications, (ii) mobile communication patterns informed by population density, and (iii) extension of wireless network infrastructures.

Geography: Previous studies have explored the geographical aspects of human social interactions in mobile networks [14, 6, 3]. These studies predominantly use traces from European countries to analyze a variety of factors such as the relationship between call duration and frequency with physical distance.

Population density: Prior work on mobile analysis informed by population density focuses on human behavioral patterns in rural and urban areas [5, 8, 9]. In contrast, we study urban and rural mobile usage to identify differences in call distance and duration. Our findings make a strong case for locality of interest in rural cellular communications in Cote d'Ivoire.

Extension of infrastructure: Several works have focused on extending limited existing communication infrastructure for cost-effective development [16, 10]. By understanding how mobile communication flows with respect to population density, we begin to understand the feasibility of a similar connectivity solution for rural Cote d'Ivoire.

3 METHODOLOGY
We explore the differences in communication patterns between three categories of antennas: Urban, Suburban and Rural. In this section we start by describing the datasets as provided by Orange. We then give details about our antenna categorization. Finally, we talk about our model for egocentric social graph analysis.

3.1 Dataset

Our analysis is based on cellular network traces provided by Orange in Cote d'Ivoire for the D4D Challenge. The datasets were collected over the course of 150 days between December 1, 2011 and April 28, 2012. For privacy purposes, customer identifiers are anonymized. Subsequent processing was performed to blur location data for antennas and generate random user identifiers for each of the four released datasets. To assure homogeneity, the data only records users who were subscribed to the network for the entire capture period. Incoming and outgoing calls associated with the same session have been combined and counted as a single call. We complement this information with a dataset from AfriPop[1] that provides high resolution population density information for Cote d'Ivoire. A detailed description of each of these datasets follows.

We utilize the population density information contained in the AfriPop data set and use Quantum GIS[2] to project this data as a raster layer. The AfriPop data includes population density information formatted as the number of people per 100 square meters. Since 100 m2 is too high of a resolution with respect to typical cellular antenna coverage, we re-sample this density data at a lower resolution creating a grid of 2 km squares where the population density assigned to each square is the mean density value of the AfriPop data bounded by this square. Each square is then assigned one of the population density categories using the OECD typology. Our grid assigns population density at a resolution suitable for associating antennas with the underlying population statistics.

4 NETWORK ANALYSIS RESULTS

We begin our analysis by mapping antennas to population density and discussing general trends in antenna utilization. We then explore trends related to population density; we expect to see differences in call duration and call frequency based on the population density of the sender's geographical area. Finally, we analyze how the inter-antenna distance impacts the call duration and call frequency.

4.1 Antenna Activity

Our analysis begins by associating antennas with their geographic location and population density. The resultant mapping of antennas to location is shown in Figure 1. The figure presents average population density per sub-prefecture with overlaid antennas. It is evident that antennas are densely clustered in urban locations while more sparsely located in rural regions. We also find that high activity antennas are often located along major transportation corridors.

We explore the relationship between the population density of a sending antenna and the average number of outbound calls associated with the antenna. Because of the predominant use of "Calling Party Pays" (CPP) policy in sub-Saharan Africa, we focus on the number of outbound calls rather than incoming calls [7]. Due to the CPP policy, we anticipate a larger mean number of outbound calls from antennas in high population density areas, which coincide with Cote d'Ivoire's financial district and center of commerce.

[1] http://www.afripop.org

[2] http://www.qgis.org/

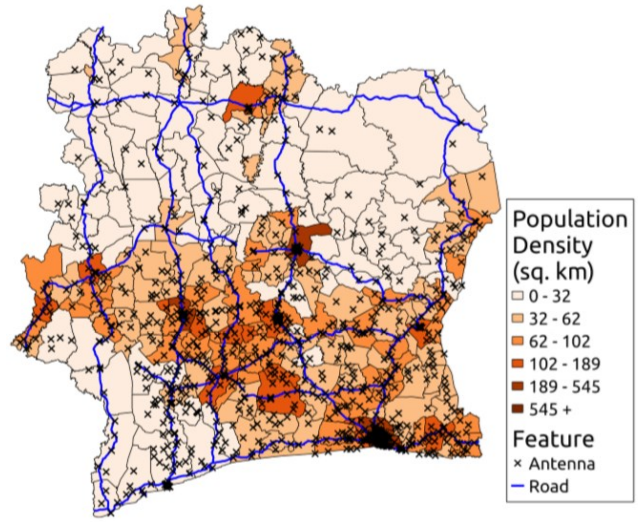

Figure 1: Cellular Antennas in Ivory Coast

Table 1: Antenna Density Classifications

Classification	Antenna Count	Source Calls
Rural	528	146,481,488
Suburban	90	21,529,115
Urban	598	331,630,147
Unknown	15	65,393,926

In Table 1 we show the number of antennas that fall into each of the classifications as well as the total number of calls originated from each antenna type. As expected, while the number of Rural and Urban antennas is almost the same, the amount of calls originated by Urban antennas is more than twice as large as those originated by Rural antennas. Very few antennas and calls are classified as Suburban. This is consistent with the fact that the subset of Suburban population is very small in comparison with Urban and Rural. In the rest of this paper we will focus our evaluation on activity associated with Rural and Urban antennas.

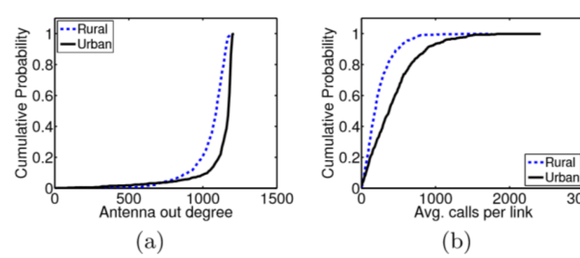

Figure 2: (a) Antenna out degree and (b) antenna out weight

We finalize our antenna activity evaluation by examining the outbound communication trends per antenna pair in Urban and Rural areas. Again, we focus on call originators. In particular, we evaluate over the entire period: (i) the outbound degree of each antenna, meaning the number of connections each antenna establishes with other antennas, and (ii) the sum outbound weight of each antenna expressed as the sum of the number of calls originated on each outbound link. We present our results in Figure 2. Figure 2(a) plots the CDF of out degree of antennas in Urban and Rural areas. As we can see, the mean outbound degree for Rural antennas is lower than that for Urban. This indicates that Rural antennas tend to communicate with fewer antennas than Urban. We then evaluate the average strength of antenna to antenna links by examining the average weight of the outgoing links associated with a given antenna; link weights are assigned according to the number of originated calls. Figure 2(b) plots a CDF of average link weight for Rural and Urban antennas. The average weight for Rural is half of that for Urban, which implies that on average more calls originate from Urban than from Rural antennas.

4.2 Call Typology Classifications

We investigate the potential correlation between population density and calling patterns by associating antennas with the corresponding local population density. This process yields antennas denoted as Rural (R), Suburban (S), Urban (U), or Unknown (for the antennas that have no geographic location). We process the Antenna-to-Antenna set to classify call records by each typology source and destination pair in order to investigate potential communication patterns. In this analysis we do not consider records for antennas without geographic data or records without valid antenna IDs.

We start by analyzing the distribution of antenna pairs in four categories: U-U, R-R, U-R and Other. Note that in this classification we do not consider directionality. The Other category contains antenna pairs featuring Suburban antennas as well as those that are unclassified. As seen in Figure 3, the majority of connections are U-U. This is followed by 20% of connections classified as R-R. Mixed links of R-U account for 18% of the total. Next, we search for differences in mean call duration across the connection classifications and show results in Figure 4(a). We find that the two call classifications with the longest mean call duration are Urban to Rural and Rural to Urban. An observable phenomenon is that calls confined to the same source and destination density type are noticeably shorter on average compared to calls between mixed types. Given our prior finding of the relationship between call distance and average duration, we posit that the majority of calls that do not cross classification boundaries are confined to a smaller geographic region.

For instance, we believe Urban to Urban calls are more likely to be sourced from and destined for the same urban area. Lastly, we observe that calls originating from Urban antennas generally have a longer duration for any destination type. This is likely due to the common CPP policy and higher buying power of individuals who reside in urban areas.

This trend leads us to look at the average distance between connecting antennas associated with each connection density classification type, shown in Figure 4(b). The longest average distance between connecting antennas occurs in the Rural to Urban and Urban to Rural cases. The shortest average distance occurs between similar source/destination connections. In the case of Urban to Urban this is likely due to the fact that there are only two Urban areas, which are likely closer to one another in comparison with an arbitrary Urban to Rural case. In contrast, the vast majority of the country is Rural, thus statistically Rural to Rural communications can cover long distances. However, the Rural to Rural patterns presented in Figure 4(b) indicate that people in Rural areas who call one another tend to be in close proximity. This indicates high locality of interest in Rural to Rural cellular communications in Cote d'Ivoire.

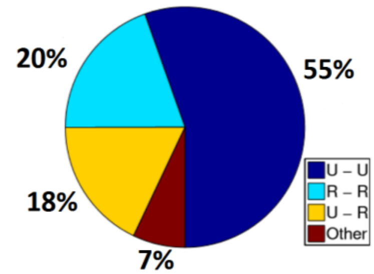

Figure 3: Classification of communication between antenna pairs

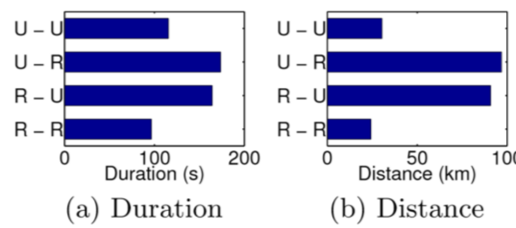

Figure 4: Mean call duration and mean call distance for connections of different types

To further explore this locality of interest, we associate call patterns and population density for calls that have the same source and destination antenna ID. We find that 57% of all Rural to Rural calls are sourced from and destined to the same antenna. We posit that this is due to fewer available antennas in predominantly rural areas. Furthermore, the coverage area of a single antenna in rural settings is typically larger (up to 35 km), which means that a higher proportion of local users are associated with the same antenna. Nevertheless, this high percentage of same antenna calls confirms that cellular communications in Rural areas are very local. In contrast, Urban connections sourced from and destined to the same antenna represent 23% of all Urban to Urban calls. We believe that the higher density and smaller cell range of Urban antennas provides more diverse antenna association possibilities for users.

4.3 Transportation Infrastructure

A close analysis of Figure 1 shows that a large fraction of antennas in areas of low population density are situated in close proximity to major transportation corridors in Cote d'Ivoire. In our current antenna typology, these antennas are categorized as Rural antennas. However, we expect that the usage patterns of antennas associated with transportation corridors will differ significantly than those located in Rural residential areas. Thus, we divide the Rural category into two subcategories: Transportation and Rural-residential, where an antenna is labeled Transportation if it is within 5 km of a highway. As a result we find that 51.7% of the antennas that were originally classified as Rural are associated with road infrastructure. We acknowledge that Transportation antennas can be used by both travelers and Rural residents, thus we feature both Transportation and Rural-residential antennas in the Rural antenna type in all other results sections.

Based on our new classification, we evaluate communication patterns in terms of three indicators: (i) call duration, (ii) number of calls and (iii) percentage of calls where both the originator and the terminator are associated with the same antenna. We find that the average call duration for Rural-residential to Rural-residential calls slightly decreased from 95.5 seconds to 84.8 seconds. More drastic was the change in the same antenna calls and the average distance between calling parties. While previously the same antenna Rural calls were 57%, in the case of Rural-residential communication the same antenna calls increases to 70%. Accordingly, the average distance between Rural to Rural calls decreased from 24.2 km to 13.4 km after the Transportation antennas were removed from the Rural typology. Once again our results confirm the strong locality of interest in cellular communications in rural Cote d'Ivoire.

5 DISCUSSION AND CONCLUSION

We show that urban and rural usage patterns in Cote d'Ivoire differ significantly. While the total number of rural and urban antennas is comparable, rural antennas tend to connect with fewer antennas than do urban antennas. Furthermore, a rural antenna originates fewer calls on average than an urban antenna. Additionally, locality of interest in rural communication is much higher than in urban communications: 70% of the Rural-residential calls occur in the vicinity of the same antenna. This means that people in rural residential areas tend to call other rural residents in close physical proximity. Although global connectivity is the ultimate goal, this finding makes a strong case for the feasibility of low-cost local cellular communication solutions [16, 10]. We expect that local cellular communication solutions would increase adoption rates of smart mobile technologies by making rural users more visible in the deployment of infrastructure. By holding rural users at the center of design, we believe that localized solutions will reduce the resource burdens of accessing cellular infrastructure for communities with low population density. However, the actual applicability of these solutions depends not only on availability of low-cost power-saving technology but also on licensing. The current licensing schemes are extremely conservative, costly and oblivious to return of investment. We argue that these licensing aspects are the main factor preventing entrepreneurs from deployment in rural remote areas, in spite of perceived national deployment of cellular infrastructure.

6 ACKNOWLEDGMENTS

We are grateful to Orange for providing these datasets as part of the D4D Challenge. This work was supported in part through NSF Network Science and Engineering (NetSE) award CNS-1064821.

7 REFERENCES

[1] European Commission urban-rural typology. http://epp.eurostat.ec.europa.eu/statistics_explained/index.php/Urban-rural_typology. Accessed: February 2013.

[2] A.Botha and L. Butgereit. Dr Math: A Mobile Scaffolding Environment. International Journal of Mobile and Blended Learning, 4(2):15–29, April 2012.

[3] V. Blondel, G. Krings, and I. Thomas. Regions and borders of mobile telephony in Belgium and in the Brussels metropolitan zone. In The e-journal for academic research on Brussels, Issue 42, October 2010.

[4] J. Burrell. Invisible users: Youth in the Internet cafes of urban Ghana. The MIT Press, May 2012.

[5] F. Calabrese and C. Ratti. Real Time Rome. In Networks and Communication Studies – Official journal of the IGU's Geography of Information Society Commission, 20:3, 247-258, 2006.

[6] F. Calabrese, Z. Smoreda, V. Blondel, and C. Ratti. Interplay between telecommunications and face-to-face interactions: A study using mobile phone data. In PLoS ONE 6(7), 2010.

[7] J. Donner. The rules of beeping: exchanging messages using missed calls on mobile phones in sub-Saharan Africa. International Communications Association, May 2005.

[8] N. Eagle, Y. de Montjoye, and L. Bettencourt. Community computing: Comparisons between rural and urban societies using mobile phone data. In International Conference on Computational Science and Engineering, volume 4, pages 144–150, 2009.

[9] M. C. Gonzalez, C. A. Hidalgo, and A.L. Barabasi. Understanding individual human mobility patterns. In Nature, Vol. 453, No. 7196., pp. 779-782, 05 June 2008.

[10] K. Heimerl, K. Ali, J. Blumenstock, B. Gawalt, and E. Brewer. Expanding Rural Cellular Networks with Virtual Coverage. In NSDI, April 2013.

[11] International Telecommunications Union. The World in 2011; Facts and Figures. http://http://www.itu.int/en/ITU-D/Statistics/Documents/facts/ICTFactsFigures2013.pdf, 2011.

[12] M. Kenney and B. Pon. Structuring the smartphone industry: is the mobile internet os platform the key? Journal of Industry, Competition and Trade, 11(3):239–261, 2011.

[13] I. Mbiti and D. N. Weil. Mobile Banking: The Impact of M-Pesa in Kenya. Working Paper 17129, National Bureau of Economic Research, June 2011.

[14] J. Onnela, S. Arbesman, M. Gonzalez, and a. N. C. A. Barabasi. Geographic constraints on social network groups. In PLoS ONE 6(4), 2011.

[15] M. Paik, J. Chen, and L. Subramanian. Epothecary: cost-effective drug pedigree tracking and authentication using mobile phones. In ACM MobiHeld, August 2009.

[16] M. Zheleva, A. Paul, D. L. Johnson, and E. Belding. Kwiizya: Local cellular network services in remote areas. In ACM MobiSys, July 2013.