BEHAVIORAL INFERENCE ACROSS CULTURES: USING TELEPHONES AS A CULTURAL LENS

NATHAN EAGLE
nathan@mit.edu

Abstract. The majority of humans today carry mobile telephones. These phones automatically capture behavioral data from virtually every human society, stored in service provider databases around the world. This article discusses the different types of data captured and how they can be used to provide insight into human cultures. Examples are provided from a variety of cultures and hundreds of millions of individuals, illustrating how phones can be used as a cultural lens, improving our understanding of a particular cultures pace of life, reactions to outlier events, and social networks.

Keywords: *data-mining, telecommunications, social networks, call graphs, privacy*

1. Introduction

Within the last decade, the amount of behavioral data about human societies has grown by many orders of magnitude. While new methods of quantifying continuous email and other online interactions have resulted in part of this wealth of information [1], much of the data is not limited to human behavior occurring online. With the ubiquity of telecommunication systems, credit cards, RFIDs, and a growing suite of additional tracking technologies, it is rapidly becoming possible to quantify detailed cultural dynamics of real-world complex social systems.

Today, most people on Earth carry mobile telephones. As a result, mobile phone service providers have access to behavioral and social network data for over 3 billion people, the majority of whom live in the developing world. Indeed, there are many nations whose entire population is accounted for in these databases. This article discusses the potential of these data using samples from several diverse cultures ranging from hundreds of students from American and Kenyan universities, to thousands of office workers in Helsinki, and millions of Rwandan and British telephone subscribers. Three examples will be used to highlight how these data can enable cultural insight: the pace of life, reactions to outlier events, and social support.

Date: June 29, 2008
to appear: *IEEE Intelligent Systems*, August 2008 .

Mobile phones are ideally suited to provide insight into social behaviors. They are inconspicuous, typically carried by the majority of individuals within a population, and today have the computational horsepower of the personal computers of only a decade ago. In particular, it is the passive sensing capabilities of current phones that make them such an important tool to study human populations. For example, this data includes continuous recordings of a) location, b) proximity, and c) communication patterns.

2. Data Types

Location inference has traditionally been done using a landline phones area code or information about the cellular towers near a mobile phone. GPS chip-sets have become increasingly common in high-end phones, however, enabling services that can ping the location of a particular phone with an accuracy of up to 10 meters. Through the analysis of tower transitions, or simply the latitude/longitude coordinates from the GPS, it is possible to infer the amount and speed of travel for an individual phone user. Aggregate data about travel patterns is a key variable for quantifying the pace of life within different cultures, epidemiological models, and next-generation applications such as anticipatory computing.

While location is obviously an important feature in this data, there is an additional sensor on modern phones of equal importance. Bluetooth, a short-range RF protocol initially designed as a cable replacement, has been installed on over 1 billion devices worldwide. When a Bluetooth device is visible, any other device conducting a scan within 3-5 meters will be able to detect its unique MAC address. By combining this proximity information with location, time, and date, it becomes possible to infer the nature of relationships between individuals. For example, regular proximity detected by a coffee machine in the afternoons represents a very different relationship than proximity detected downtown late on Saturday night [2].

Perhaps a more obvious method of extracting social information from phones is through the analysis of communication patterns, via calls or text messages. By looking at trends in the duration and frequency of calls, it becomes possible to learn about the social networks of individuals. This information about communication patterns can be coupled with user surveys to gain further insights. For example in our Helsinki study the phones were programmed to launch a survey question about the type of relationship the user had with the individual he or she just called. Since it is not feasible to conduct this type of survey for an entire population, another type of social data is the adoption of service (tariff) plans and other telecommunication products, which may be thought of as the spread of a social contagion. In our larger datasets (consisting of more than 500,000 people) that there are particular individuals who hold influence over others in their peer group; when they adopt a particular product or service, the individuals whom they call subsequently adopt the product as well. Through the analysis of the diffusion of these social contagions over a call graph it may become possible to learn more about the social dynamics inherent within a population.

An important consideration for any study of this kind is that the telecommunication industry does not generally seek consent from their subscribers to participate in this research. While privacy laws throughout the world differ regarding call log record regulations, it is the researchers responsibility to ensure that the data is appropriately hashed such that it is impossible to gain access to any individuals actual phone logs or personal information an increasingly challenging, yet critical requirement.

3. Pace of Life

We have shown in prior work that it is possible to infer an individuals daily commute to work, as well as the amount of time spent at work, at home and traveling [3]. Now with access to several months of continuous cellular tower transition and communication data from thousands of individuals across Africa, Europe and the United States, it becomes possible to compare aggregate movement and communication metrics. While the initial research involves comparing these different cultures and tracking how they change over time, it is also possible to quantify generalities about universal human activities such as urbanization.

The relationship between cities and the pace of life has recently been the subject of much academic research [4]. While it is possible to compare the communication behaviors between residents of rural and urban communities, causality remains elusive. Rural individuals who already have a fast pace of life and a large, active social network may gravitate towards a city, rather than the city actively increasing the pace of life of its inhabitants. However, with access to several years of behavioral data for hundreds of thousands of people living in Rwanda, it may now be possible to disentangle the effect cities have on the social networks and lifestyle of its inhabitants. By identifying subscribers who have relocated from rural Rwanda to Kigali, Rwandas capital, a comparison can be made between their routine behavior before and after the move. The hypothesis is that the move from a rural to urban area will be correlated with an increase in social contacts, speed of travel, and the individuals influence over his contacts, as measured by the diffusion of products and services.

4. Reactions to Outlier Events

By continuously recording behavior over an extended period of time, it will become possible to quantify peoples reactions to unanticipated events such as an election or an earthquake. An outlier event is characterized as one in which the majority of individuals are removed from their typical routines and behave in a way that was unpredicted by our behavioral models. Two such events occurred while collecting data from the students at MIT and the University of Nairobi.

The original Reality Mining study ran from August 2004 May 2005 and involved 100 MIT students and staff living in the greater Boston area. Data about location, physical proximity, and communication was continually logged for the duration of the study [5]. During the autumn of 2004, while the data collection was taking place, the Boston Red Sox won their first World Series. At this point, the behavior of the subjects was no longer governed by previous routines instead, many subjects (along with tens of thousands of other fans) converged in central Boston for a spontaneous rally. Urban planners working for the city of Boston were interested in this data not only to try to understand what sparked the rally in this particular location, but also to learn how the participating individuals used the urban resources to disperse once the rally ended. Even with only coarse cellular tower transitions from a sample of the individuals in the rally, estimates could be made about how many people walked, biked, drove, and used public transportation to and from the area.

A follow-up Reality Mining study ran from September 2006 May 2008 involving very different subject demographics. This study took place in Kenya and involved not only university students, but day laborers, taxi drivers, and even individuals who had never before owned a phone. The data collection overlapped with a major presidential election at the end of December 2007 and the subsequent collapse of the Kenyan government in early 2008 as the country veered towards civil war. With the spread of violence throughout the country, the data collected from these subjects will enable detailed insight into how Kenyans movement and communication patterns changed in response to the civil unrest, with the end goal of better informing the design of new methods for limiting future violence.

5. Social Support

Communication patterns, from individual behavior to national call graphs, can lend a significant amount of insight into culture. This section will discuss how to identify different types of students based on the expansion rate of his/her egocentric social network. Additionally, we demonstrate that it is not the number of phone calls that is an indicator of socioeconomic status, but rather the diversity of calls (defined by calls to many different regions). Finally, a new type of mobile phone data only recently being collected in locations such as Kenya is introduced: financial transactions.

It has been shown that an individuals social network can be broken down into hierarchies ranging from the support group (typically the core group of 3-5 people relied on in times of severe financial or emotional distress), the sympathy group (12-20 friends or colleagues), and the cognitive group (the 150 people with whom individuals maintain a rare but coherent personal relationship) [6]. While it is impossible to identify all of an individuals relationships, telephone communication provides a significant amount of information about an individuals social network. Normalizing for the cost of communication and income levels, one interesting research question is a structural comparison of the egocentric social networks in the US, the UK, Finland, Kenya and Rwanda. Even our study focusing on a single American university, for example, revealed surprising contrasts between the different types of incoming students. When most students arrive on campus for the first time, their social network rapidly expands for several months until it reaches an approximately steady state, at which point they have found their group of friends. The number of unique phone numbers the students dialed consistently throughout the year was used as a proxy for this network expansion. However, the social network of a particular group of students at MIT did not reach a steady state at the same time as the other students. During the first two months the thirty incoming business school students egocentric social networks expanded at the same rate as all other incoming students, yet their network never stopped growing at any point during the 9-month study. Business school students have a culture of networking and place high value on the contacts they make in school; these types of values are immediately apparent in their call logs.

These American business school students seem to have adopted a positive cultural behavior: in our analysis of communication logs from the UK, individuals who communicate with a variety of different people have a greater socioeconomic status. While it isnt possible to establish causality between this behavior and socioeconomic status, there is a significant ($R = -0.75, p < .001$) correlation between a regions communication diversity and its index of deprivation (the metric for socioeconomic status of the UK Civil Service). It certainly seems plausible that some cultures encourage interactions with others while other groups prefer to remain insular. In the US this type of culture was shown above to be associated with business school students and in the UK the culture appears to be associated with individuals of higher socioeconomic status. The question whether this result is universal across countries is actively being pursued. Although diverse communities (both insular and cosmopolitan) in the call graph have been discovered within Rwanda, the appropriate regional Rwandan socioeconomic data is still being collected.

In Kenya, mobile phones can be used to infer relationships not only from communication but also financial support, as it is now possible to send money to and from any mobile phone within the network. There are currently millions of Kenyans transferring airtime and currency using their mobile phones. This financial network contains distinct, but overlapping information when juxtaposed with the standard call graphs. A collaboration with the Kenyan mobile phone operator is being formed to learn more about the cultural dynamics of the thousands of Kenyan tribes, each with their own local language and geographic location and attempt to understand how the tribes are connected to each other both through communication as well as through financial exchange.

6. Concluding remarks

Whether it is to study the differences between groups of local students, quantify the behavioral effects of urbanization, or infer the relationships between ethnic tribes across a continent, the data left in the wake of mobile phones can provide invaluable information

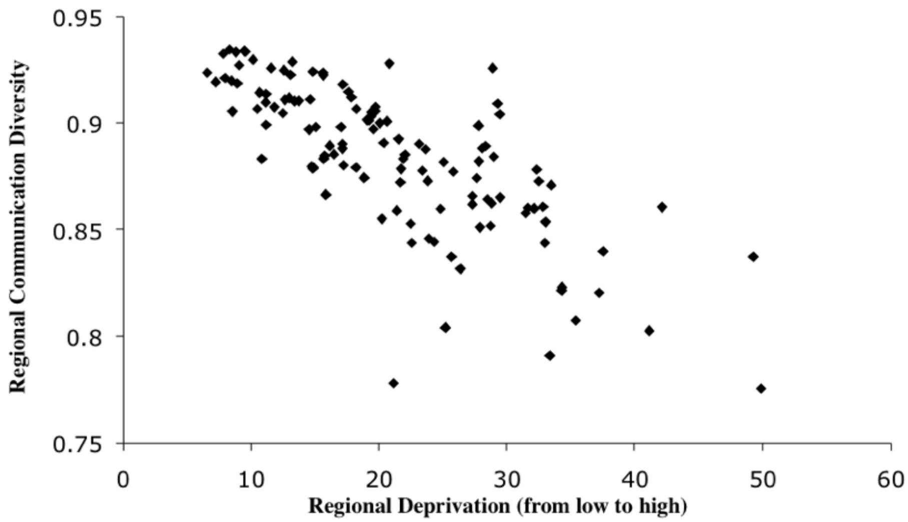

Figure 1. A preliminary plot of regional communication diversity and the corresponding index of deprivation (a socioeconomic status measurement that is a combination of metrics such as average income levels, access to healthcare, and education). This data came from the UK call graph consisting of 250 million hashed phone numbers and 12 billion phone calls. It can be seen that the regions with a culture of creating diverse social ties tend to be the least deprived. Diversity of communication behavior has a significant correlation of r=-.75 with the regional index of deprivation.

about the cultural dynamics of our species. While some may argue that this type of pervasive behavioral data should not be collected, it is a fact of life in the 21st century the data discussed here will continue to be aggregated by the hundreds of mobile phone service providers throughout the world, whether or not it is shared with researchers. Therefore while academics must remain cognizant of the privacy issues surrounding the analysis of personal information, society has much to gain from these studies and their potential for use in solving social problems ranging from disease outbreaks to urban planning. To achieve these goals, new tools will be needed to grapple with data sets that are many orders of magnitude larger than have previously existed.

References

[1] G. Kossinets and D. J. Watts, Empirical Analysis of an Evolving Social Network, Science 311, 2006, pp. 88-90.
[2] N. Eagle, A. Pentland, and D. Lazer, "Inferring Social Network Structure using Mobile Phone Data," 2007. (in submission).

[3] N. Eagle and A. Pentland, "Eigenbehaviors: Identifying Structure in Routine," Behavioral Ecology and Sociobiology, 2008. (in press)
[4] L. M. A. Bettencourt, J. Lobo, D. Helbing, C. Kuhnert, and G. B. West, Growth, innovation, scaling, and the pace of life in cities, PNAS 104, 2007, pp. 7301-7306.
[5] N. Eagle and A. Pentland, "Reality Mining: Sensing Complex Social Systems," Personal and Ubiquitous Computing, vol 10, 4, 2006. pp. 255-268.
[6] W.-X. Zhou, D. Sornette, R.A. Hill and R.I.M. Dunbar, Discrete Hierarchical Organization of Social Group Sizes, Proc. Royal Soc. 272, 2005, pp. 439-444.

NATHAN@MIT.EDU