INSEAD

The Business School for the World®

WORLD ECONOMIC FORUM

COMMITTED TO IMPROVING THE STATE OF THE WORLD

# The Global Information Technology Report 2008–2009

## Mobility in a Networked World

Soumitra Dutta
Irene Mia

CHAPTER 1.6

# Reality Mining of Mobile Communications: Toward a New Deal on Data

**ALEX PENTLAND,** Massachusetts Institute of Technology (MIT)

> Within just a few years "people data" will be 90% of the world's collective data.
>
> *—Jeff Nick, CTO of EMC, personal communication*

> We have enough water, enough food, enough money; we have enough of everything except the ability to agree and move forward.
>
> *—Abdul Kalam, former President of India, personal communication*

Around the world, many of us live our lives in digital networks. We wake up in the morning, check our email, make a quick phone call, commute to work, and buy lunch. Many of these transactions leave digital breadcrumbs—tiny records of our daily experiences.[1] *Reality mining,* which pulls together these crumbs using statistical analysis and machine learning methods, offers increasingly extensive information about our lives, both individually and collectively. This technology's potential to transform our understanding of ourselves, our organizations, and our society is the reason that MIT's *Technology Review* recently identified reality mining as one of "10 Emerging Technologies That Will Change the World."[2]



## An emerging—and truly global—nervous system

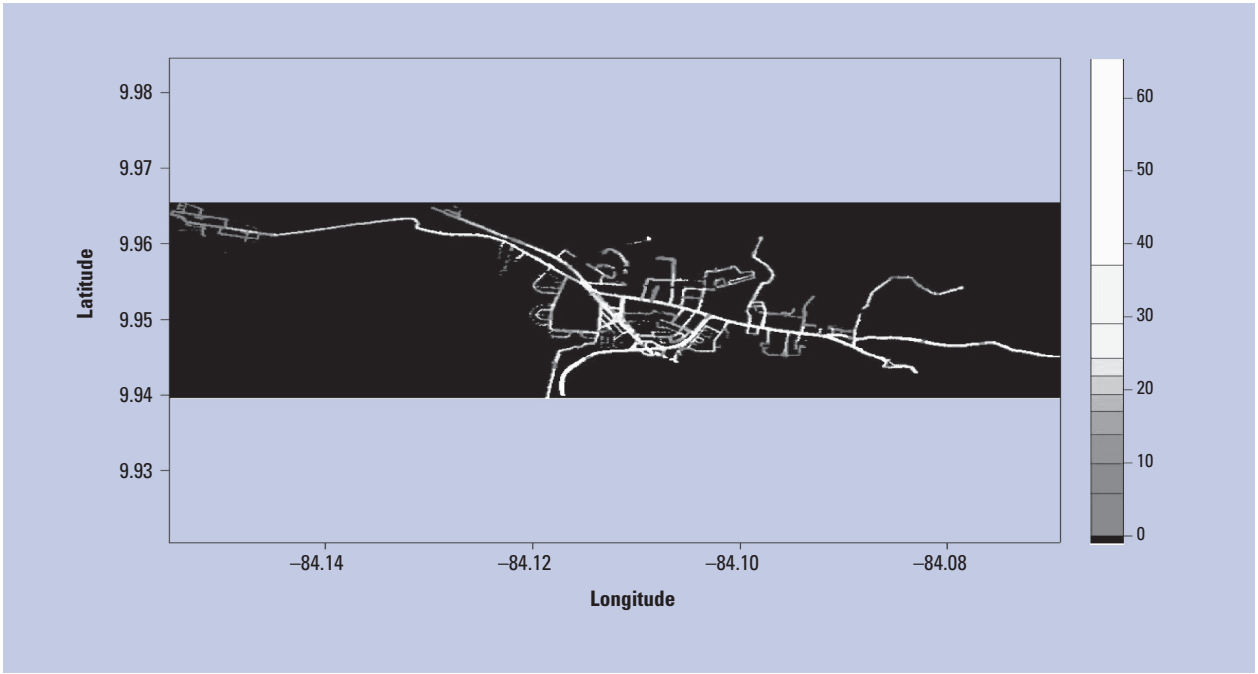
The ability to understand the patterns of human life by analyzing the digital traces that we leave behind will transform poor nations even more than rich nations. Ten years ago, half of humanity had never made a phone call and only 20 percent had regular access to communications. Today almost 4 billion people have a mobile telephone, and because remanufactured mobile phones cost US$10 in the developing world and incoming messages are free, almost every stratum of society is now connected.[3] For the first time, the majority of humanity is linked and has a voice.

The most important changes may come, however, from the fact that these same mobile phones can be used for reality mining. Their reality mining functionality is mostly latent at this point, but already these devices are being used to measure population flows into cities and slums, to map the movement of populations during emergencies, to identify neighborhoods where social services are inadequate (see Figure 4), and to manage automobile traffic congestion (see Figure 1).

The reality mining functionality of mobile phone networks is what governments in countries such as India are using to track terrorists, and they claim that the vast

This article builds on papers co-authored with David Lazer at the J. F. Kennedy School of Government at Harvard University and with my current and former students, as well as on continual interactions and conversations with my colleagues at Sense Networks and at the MIT Media Laboratory.

**Figure 1: Traffic congestion predicted using mobile phone GPS data**




Source: Dong, 2006.



majority of captured terrorists have been identified through mobile phone transactions. The ability of mobile phone networks to identify unusual patterns of movement and communication are also how public health officials and disaster relief teams are scanning for outbreaks of diseases such as severe acute respiratory syndrome (SARS) and emergencies such as tidal waves.

It seems that the human race suddenly has the beginnings of a working nervous system. Like some world-spanning living organism, automobile traffic systems, security sensors, and especially mobile telephone networks are all becoming intelligent, reactive systems with sensors serving as their eyes and ears. Moreover, the evolution of this nervous system will continue at a quickening speed because of the exponential progress in computing and communication technologies as well as basic economics. Networks will become faster, devices will have more sensors, and techniques for modeling human behavior will become more accurate and detailed.

## From individuals to societies

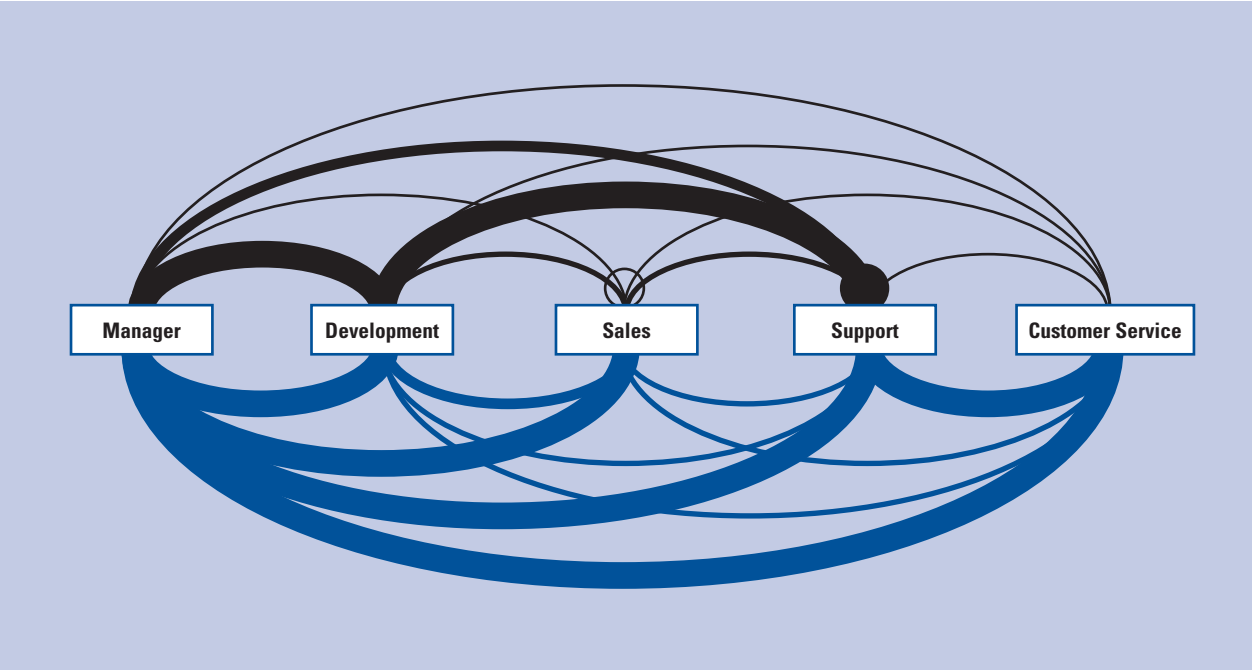
To date, the vast majority of research on the human condition has relied on single-shot, self-report data on relationships: a yearly census, public polls, focus groups, and the like. Reality mining offers a remarkable, second-by-second picture of group interactions over extended periods of time, providing dynamic structural and content information.

Perhaps the simplest example of reality mining is the analysis of automobile traffic congestion by using the global positioning system (GPS) data collected from the mobile telephones carried by the automobile drivers. These data provide minute-by-minute updates on traffic flow, allowing for more accurate predictions of driving time. Congestion patterns can be predicted days in advance, and traffic jams detected hours before they become serious, as illustrated by the "congestion map" of city streets shown in Figure 1.

Similar to using reality mining to understand traffic within a city, we can also use reality mining of mobile phone GPS data, call logs, and email records to better understand the "traffic" within an organization. Analysis of these digital traces allows a detailed picture of face-to-face, voice, and digital communication patterns. These patterns, in turn, allow us a new level of insight into the problems of industry and government, including building customer relationships, resource management, transportation, and public health. Figures 2 and 3 show two recent but very different applications of reality mining of people data.

Figure 2 shows the pattern of face-to-face and email communication within a German bank. The thickness of the top arcs (in black) shows the amount of face-to-face communication, while the thickness of the bottom arcs (in blue) shows the amount of email communication. In our studies of many different types of companies, we have found that the tradeoff between face-to-face and email communication is a critical

**Figure 2: Patterns of communication within departments of a bank**




Source: Pentland, 2008.



predictor of both productivity and job satisfaction. In fact, variations in these patterns can account for more than 30 percent of the differences in productivity between different parts of the same company! Patterns of *change* within these communication networks are also a critical predictor of creative productivity, accounting for up to 40 percent of differences in creative output.[4]

Figure 3 shows a 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 health care, and education. These data, drawn from town councils across the United Kingdom, show that it is possible to use patterns of communication to identify "information ghettos" that have serious social problems. This capability can allow government services to be far more responsive to citizen needs than is possible by using census or survey data.

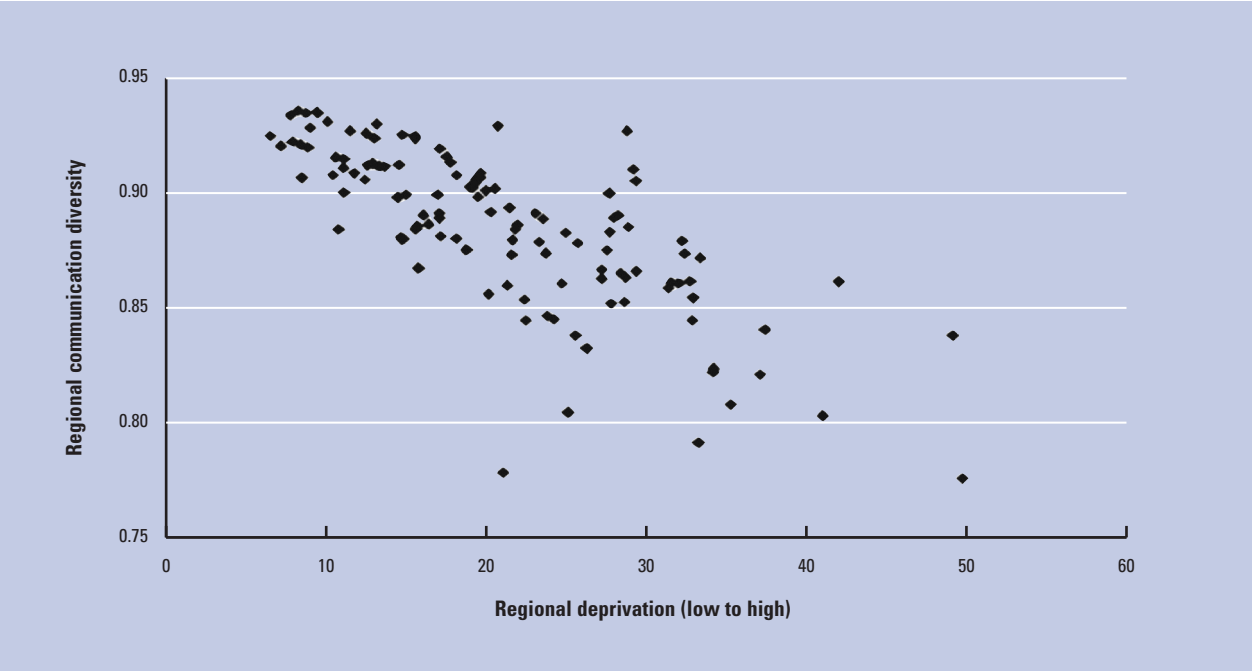
In short, reality mining is beginning to provide the capacity to collect and analyze data about people with a breadth and depth that was previously inconceivable. Current work using reality mining techniques is underway in a variety of applications. For example, reality mining studies of voice, communications, and mobility patterns have already demonstrated the ability to screen for depression, infer quality-of-life metrics, and develop financial indexes for individual neighborhoods.[5]

## Beyond demographics

Advertising, commercial development, and government services all currently rely on demographic data to guide them. Such data can quickly become out of date and, of course, good demographic data simply do not exist in many parts of the world. More importantly, however, we have found that demographics are a relatively poor predictor of behavior, and behavior is what we really need to understand in order to reach the right people.

The fact that mobile phones have GPS means that we can leap beyond demographics directly to measuring behavior. Where do people eat? Work? Hang out? How does word of mouth spread? Analysis of travel patterns using mobile phone GPS data, for instance, allows discovery of the independent subgroups within a city. Figure 4a shows movement patterns with popular "hang outs" coded by the different subgroups that populate these destinations, where the subgroups are defined by both their demographics and, more importantly, by their *behaviors.* Figure 4b shows that the mixing between these different behavior groups is surprisingly small. Knowledge of the "hang outs" of different subgroups and the mixing among groups can provide great improvements in advertising, commercial development, public education, and policy interventions.

**Figure 3: Communication patterns vs. deprivation**

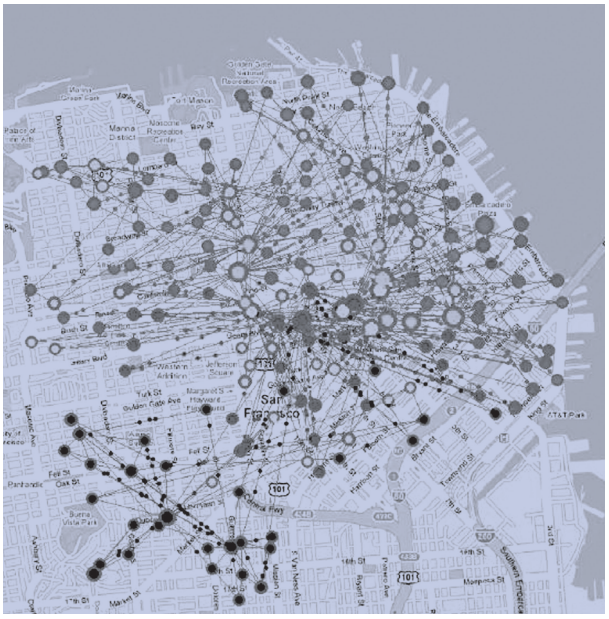



Source: Eagle, 2009.



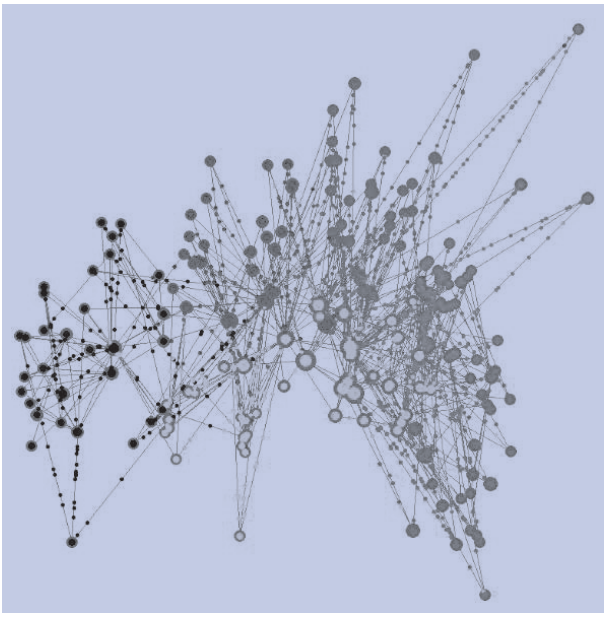
**Figure 4: Reality mining of data from GPS mobile phones**

4a: Patterns of human movement in San Francisco



Source: Sense Networks, 2008.

4b: Limited mixing among people with different behavior patterns



Source: Sense Networks, 2008.

### Usefulness for society

Reality mining of behavior data is just beginning. For instance, the correlation of behavior data with medication data from millions of people could make drug therapies more effective and help medical professionals detect drug interactions more quickly. If behavior data were correlated with medical conditions, the data could illuminate the etiology and preconditions of disease far more powerfully than is possible today and, further, serve as an early warning system for epidemic diseases such as SARS. Comparing the medical data with genomic and protonomic data from different population samples could provide a powerful method for understanding complex gene and environment interactions.

For society, the hope is that we can use this new in-depth understanding of individual behavior to increase the efficiency and responsiveness of industries and governments. For individuals, the attraction is the possibility of a world where everything is arranged for your convenience—your health checkup is magically scheduled just as you begin to get sick, the bus comes just as you get to the bus stop, and there is never a line of waiting people at city hall.

### Data ownership and privacy

Perhaps the greatest challenge posed by this new ability to sense the pulse of humanity is creating a "new deal" around questions of privacy and data ownership. Many of the network data that are available today are freely offered because the entities that control the data have difficulty extracting value from them. As we develop new analytical methods, however, this will change. Moreover, not all people who want access to the data do so for altruistic motives, and it is important to consider how to keep the individuals who generate this information safe. Advances in analysis of network data must be approached in tandem with understanding how to create value for the producers and owners of the data while at the same time protecting the public good. Clearly, our notions of privacy and ownership of data need to evolve in order to adapt to these new challenges.

This raises another important question: how do we design institutions to manage the new types of privacy issues that will emerge with these new reality mining capabilities? Digital traces of people are ubiquitously preserved within our private and public organizations—location patterns, financial transactions, public transportation, phone and Internet communications, and so on. Certainly new types of regulatory institutions are required to deal with this information, but what form should they take?

Companies will have a key role in this new deal for privacy and ownership. One suggestion is that there is an incentive system, one that gives added value to the users. Market mechanisms appear to be a particularly interesting avenue of exploration, since they may allow people to give up their data for monetary or service rewards. Ideally, this would be put into place in order to gain approval from the majority of the population to use data collected from their digital interactions.

Other important considerations revolve around data anonymity. The use of anonymous data should be enforced, and analysis at the group level should be preferred over that at the individual level. Robust models of collaboration and data sharing need to be developed; guarding both the privacy of consumers as well as corporations' legitimate competitive interests are vital here.

What must be avoided is either the retreat into secrecy, so that these data become the exclusive domain of private companies and remain inaccessible to the Common Good, or the development of a "big brother" model, with government using the data but denying the public the ability to investigate or critique its conclusions. Neither scenario will serve the long-term public interest in having a transparent and efficient government.

### The new deal on data

The first step toward open information markets is to give people ownership of their data. The simplest approach to defining what it means to "own your own data" is to go back to Old English Common Law for the three basic tenets of ownership, which are the rights of possession, use, and disposal:

1. You have a right to *possess* your data. Companies should adopt the role of a Swiss bank account for your data. You open an account (anonymously, if possible), and you can remove your data whenever you'd like.

2. You, the data owner, must have full control over the *use* of your data. If you're not happy with the way a company uses your data, you can remove it. All of it. Everything must be opt-in, and not only clearly explained in plain language, but with regular reminders that you have the option to opt out.

3. You have a right to *dispose* or *distribute* your data. If you want to destroy it or remove it and redeploy it elsewhere, it is your call.

Ownership seems to be the minimal guideline for the "new deal on data." There needs to be one more principle, however—which is to adopt policies that encourage the combination of massive amounts of anonymous data to promote the Common Good. Aggregate and anonymous location data can dramatically improve society. Patterns of how people move around can be used for early identification of infectious disease outbreaks, protection of the environment, and public

safety. It can also help us measure the effectiveness of various government programs, and improve the transparency and accountability of government and nonprofit organizations.

## Conclusions

Revolutionary new measurement tools provided by mobile telephones and other digital infrastructures are providing us with a God's eye view of ourselves. For the first time, we can precisely map the behavior of large numbers of people as they go about their daily lives.

These distributed sensor networks have given us a new, powerful way to understand and manage human groups, corporations, and entire societies. As these new abilities become refined by the use of more sophisticated statistical models and sensor capabilities, we could well see the creation of a quantitative, predictive science of human organizations and human society. At the same time, these new tools have the potential to make George Orwell's vision of an all-controlling state into a reality. What we do with this new power may turn out to be either our salvation or our destruction.

## Notes

1 See http://www.bbc.co.uk/britainfromabove/stories/visualisations/index.shtml for amazing videos of the time evolution of automobiles, airplanes, telephone calls, and shipping activity patterns.

2 *Technology Review* 2008.

3 Eagle 2009.

4 Pentland 2008.

5 Pentland 2008.

## References

Dong, W. 2006. "Influence Modeling of Complex Stochastic Processes." July. Master's Thesis in Media Arts and Sciences. Available at http://hd.media.mit.edu.

Eagle, N. 2009 (in preparation). "Mobile Phones as Social Sensors." *The Handbook of Emergent Technologies in Social Research*. New York: Oxford University Press.

Pentland, A. 2008. *Honest Signals: How They Shape our World.* Cambridge, MA: MIT Press.

Sense Networks. 2008. "Technology tab" at http://www.sensenetworks.com.

*Technology Review.* 2008. "10 Emerging Technologies That Will Change the World." Available at http://www.technologyreview.com/article/13060/.