THE DATA-DRIVEN SOCIETY

The digital traces we leave behind each day reveal more about us than we know. This could become a privacy nightmare—or it could be the foundation of a healthier, more prosperous world

By Alex "Sandy" Pentland

Purchases five cups of coffee (usually Starbucks) per day

Search engine records indicate coming down with the flu

Annual salary of $153,000

Based on social circle, likely to repay a loan

Pregnant but doesn't know it yet

BY THE MIDDLE OF THE 19TH CENTURY, RAPID URBAN GROWTH SPURRED BY THE INDUSTRIAL REVOLUTION HAD CREATED URGENT SOCIAL AND ENVIRONMENTAL PROBLEMS. CITIES RESPONDED BY BUILDING CENTRALIZED NETWORKS TO DELIVER CLEAN WATER, ENERGY AND SAFE FOOD;

Alex "Sandy" Pentland directs the M.I.T. Human Dynamics Laboratory and co-leads the World Economic Forum's big data and personal data initiatives. His next book, *Social Physics*, will be published in January 2014 by Penguin Press.

to enable commerce, facilitate transportation and maintain order; and to provide access to health care and energy. Today these century-plus-old solutions are increasingly inadequate. Many of our cities are jammed with traffic. Our political institutions are deadlocked. In addition, we face a host of new challenges—most notably, feeding and housing a population set to grow by two billion people while simultaneously preventing the worst impacts of global warming.

Such uniquely 21st-century problems demand 21st-century thinking. Yet many economists and social scientists still think about social systems using Enlightenment-era concepts such as markets and classes—simplified models that reduce societal interactions to rules or algorithms while ignoring the behavior of individual human beings. We need to go deeper, to take into account the fine-grained details of societal interactions. The tool known as big data gives us the means to do that.

Digital technology enables us to study billions of individual exchanges in which people trade ideas, money, goods or gossip. My research laboratory at the Massachusetts Institute of Technology is using computers to look at mathematical patterns among those exchanges. We are already finding that we can begin to explain phenomena—financial crashes, political upsets, flu pandemics—that were previously mysterious. Data analytics can give us stable financial systems, functioning governments, efficient and affordable health care, and more. But first we need to fully appreciate the power of big data and build a framework for its proper use. The ability to track, predict and even control the behavior of individuals and groups of people is a classic example of Promethean fire: it can be used for good or ill.

THE PREDICTIVE POWER OF DIGITAL BREAD CRUMBS

AS WE GO ABOUT our daily lives, we leave behind virtual bread crumbs—digital records of the people we call, the places we go, the things we eat and the products we buy. These bread crumbs tell a more accurate story of our lives than anything we choose to reveal about ourselves. Our Facebook status updates and tweets deliver information we choose to tell people, edited according to the standards of the day. Digital bread crumbs, in contrast, record our behavior as it actually happened.

We are social animals, and our behavior is never as unique as

---

IN BRIEF

Today's cities and governments still operate according to principles developed two centuries ago, during the industrial revolution. To address 21st-century problems such as exploding population growth and climate change, we need new thinking.

Big data can deliver that thinking. The digital bread crumbs we leave behind as we go about our daily lives—which reveal more about us than anything we choose to disclose—provide a powerful tool for tackling social problems.

Yet concerns about misuse of this information are valid. Before data mining can deliver a healthier, more prosperous society, we need a New Deal on Data that gives individuals far more control over their information than they have today.

Group Dynamics

By tracking social interactions with sensor-equipped ID badges or cell-phone data, it is possible to make group dynamics visible, enabling team members to see how they work together. In the case depicted below, the author and his team tracked the dynamics of an eight-person brainstorming group and gave members a diagram of their interactions at the end of each day. The size of the circle surrounding each member represents the amount that person communicated with the group; the width of the lines connecting individuals represents the amount those two communicated with each other. By using the diagrams to diagnose weak links, the team became more interactive and, thus, more productive.

Day 1

Interactions (*lines*)

Individuals (*nodes*)
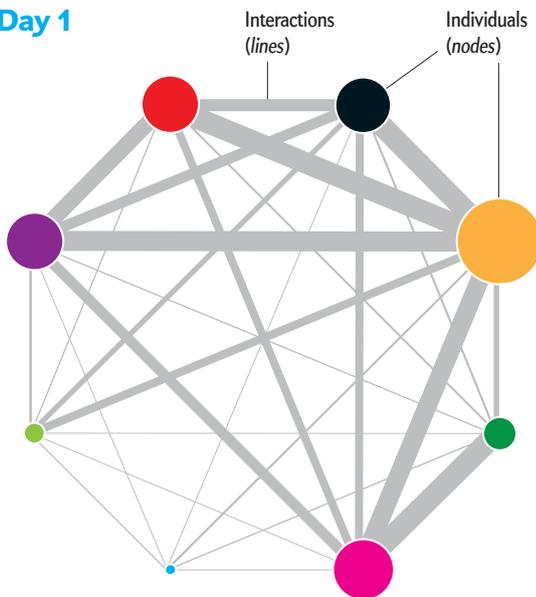
Day 7
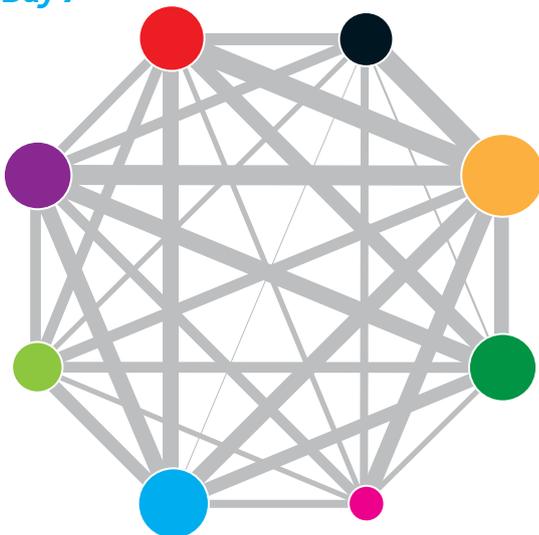
we might think. The people you call, text and spend time with—even the people you recognize around the neighborhood but have never formally met—are likely to be similar to you in all kinds of ways. My students and I can tell whether you are likely to get diabetes by examining the restaurants where you eat and the crowd you hang out with. We can use the same data to predict the sort of clothes you are inclined to buy or your propensity to pay back a loan. Because our behavior changes when we feel like we are getting sick—we go different places, buy different things, call different people and search for different terms on the Web—it is now possible, using data analytics, to make a constantly updatable map that predicts where residents of a city are most likely to come down with the flu at any given moment.

The mathematical patterns within big data that provide the most insight into the functioning of society involve the flow of ideas and information between people. We can see this flow by studying patterns of social interaction (face-to-face conversations, telephone calls, social-media messaging) and by assessing the amount of novelty and exploration in individuals' purchasing patterns (as seen in credit-card data) or movement patterns (as seen in GPS tracks). The flow of ideas is central to understanding society not only because timely information is critical to efficient systems but also because the spread and combination of ideas form the basis of innovation. Communities that are cut off from the rest of society risk becoming stagnant.

Among the most surprising findings that my students and I have discovered is that patterns of idea flow (measured by purchasing behavior, physical mobility or communications) are directly related to productivity growth and creative output. Individuals, organizations, cities, and even entire societies that engage with one another and explore outside their social group have higher productivity, greater creative output and even longer, healthier lives. We see variations on this pattern in all social species, even bees. Idea flow seems to be essential to the health of every society.

Consequently, when we analyze companies and governments, it is useful to think of them as idea machines. These machines harvest and spread ideas primarily through individual interactions. Two mathematical patterns provide evidence for healthy idea flow. The first is *engagement,* which we define as the proportion of possible person-to-person exchanges within a work group that regularly occur. The relationship between engagement and productivity is simple: high levels of engagement predict high group productivity, almost no matter what that group is working on or what kinds of personalities its members have. The second factor is *exploration*—a mathematical measure of the extent to which the members of a group bring in new ideas from outside. Exploration is a good predictor of both innovation and creative output.

In field experiments conducted at companies around the world, my students and I have measured levels of engagement and exploration by equipping employees with sociometric ID badges, electronic devices that track person-to-person interactions. We have found that increasing the amount of engagement within a group can dramatically improve productivity while simultaneously reducing stress. For instance, after learning that call centers usually schedule coffee breaks so that only one person has a break at any given time, I persuaded the manager of a Bank of America call center to schedule coffee breaks simultaneously. The goal was to promote more engagement between em-

ployees. This single change resulted in a productivity increase of $15 million a year.

We have also found that exploration—establishing new connections among people—is an excellent predictor of innovation and creative output. Rich channels of communication, particularly face-to-face interaction, matter much more than electronic communication channels. In other words, e-mail can never fully replace meetings and conversations.

We have also found that an oscillating pattern of exploration and group engagement—in which people engage the group, then go find new information, bring it back, then repeat the process—is consistently associated with greater creative output. In established research organizations, my colleagues have been able to measure this pattern in face-to-face interactions and use these measurements to accurately identify researchers' top creative days. The same approach works with virtual teams, whose members are distributed across many locations.

Similar patterns of information flow predict the productive output of entire cities and regions. Patterns of community engagement and out-of-community exploration even predict social outcomes such as life expectancy, crime rate and infant mortality. Neighborhoods that are information ghettos do as poorly as physical ghettos do, whereas neighborhoods that are engaged with one another and connected to surrounding communities tend to be more healthy and prosperous.

MAXIMIZING IDEA FLOW

USING BIG DATA to diagnose problems and predict successes is one thing. What is even more exciting is that we can use big data to *design* organizations, cities and governments that work better than the ones we have today.

The potential is easiest to see within corporations. By measuring idea flow, it is usually possible to find simple changes that improve productivity and creative output. For instance, the advertising department of a German bank had experienced serious problems launching successful new product campaigns, and they wanted to know what they were doing wrong. When we studied the problem with sociometric ID badges, we found that while groups within the organization were exchanging lots of e-mails, almost no one talked to the employees in customer service. The reason was simple: customer service was on another floor. This configuration caused huge problems. Inevitably, the advertising department would end up designing ad campaigns that customer service was unable to support. When management saw the diagram we produced depicting this broken flow of information, they immediately realized they should move customer service to the same floor as the rest of the groups. Problem solved.

Increasing engagement is not a magic bullet. In fact, increasing engagement without increasing exploration can cause problems. For instance, when postdoctoral student Yaniv Altshuler and I measured information flow within the eToro social network

HOT ZONES: By mining cell-phone data, researchers can map the places where people are most likely to catch the flu. Above, an example from an experiment on the M.I.T. campus.

For the first time in history, we can see enough about ourselves to build social systems

that work better than the ones we have always had.

of financial traders, we found that at a certain point people become so interconnected that the flow of ideas is dominated by feedback loops. Sure, everyone is trading ideas—but they are the same ideas over and over. As a result, the traders work in an echo chamber. And when feedback loops dominate within a group of traders, financial bubbles happen. This is exactly how otherwise intelligent people all became convinced that Pets.com was the stock of the century.

Fortunately, we have found that we can manage the flow of ideas between people by providing small incentives, or nudges, to individuals. Some incentives can nudge isolated people to engage more with others; still others can encourage people mired in groupthink to explore outside their current contacts. In an experiment with 2.7 million small-time, individual eToro investors, we "tuned" the network by giving traders discount coupons that encouraged them to explore the ideas of a more diverse set of other traders. As a result, the entire network remained in the healthy wisdom-of-the-crowd region. What was more remarkable is that although we applied the nudges only to a small number of traders, we were able to increase the profitability of *all* social traders by more than 6 percent.

Designing idea flows can also help solve the tragedy of the commons, in which a few people behave in such a way that everyone suffers, yet the cost to any one person is so small there is little motivation to fix the problem. An excellent example can be found in the health insurance industry. People who fail to take medicine they need, or exercise, or eat sensibly have higher

health care costs, driving up the price of health insurance for everyone. Another example is when tax collection is too centralized: local authorities have little incentive to ensure that everyone pays taxes, and as a result, tax cheating becomes common.

The usual solution is to find the offenders and offer incentives or levy penalties designed to get them to behave better. This approach is expensive and rarely works. Yet graduate student Ankur Mani and I have shown that promoting increased engagement between people can minimize these situations. The key is to provide small cash incentives to those who have the most interaction with the offenders, rewarding *them* rather than the offender for improved behavior. In real-world situations—with initiatives to encourage healthy behavior, for example, or to prompt people to save energy—we have found that this social-pressure-based approach is up to four times as efficient as traditional methods.

This same approach can be used for social mobilization—in emergencies, say, or any time a special, coordinated effort is needed to achieve some common goal. In 2009, for example, the Defense Advanced Research Projects Agency designed an experiment to celebrate the 40th anniversary of the Internet. The idea was to show how social media and the Internet could enable emergency mobilization across the U.S. DARPA offered a $40,000 prize for the team that could most quickly find 10 red balloons placed across the continental U.S. Some 4,000 teams signed up for the contest, and almost all took the simplest approach—offering a reward to anyone who reported seeing a balloon. My research group took a different tack. We split the reward money among those who used their social networks to recruit a person who later saw a balloon and those who saw a balloon themselves. This scheme, which is conceptually the same as the social-pressure approach to solving tragedies of the commons, encouraged people to use their social networks as much as possible. We won the contest by locating all 10 balloons in only nine hours.

A NEW DEAL ON DATA

TO ACHIEVE a data-driven society, we need what I have called the New Deal on Data—workable guarantees that the data needed for public goods are readily available while at the same time protecting the citizenry. The key to the New Deal is to treat personal data as an asset; individuals would have ownership rights in data that are about them. What does it mean to "own" your own data? In 2007 I suggested an analogy with the English common law tenets of possession, use and disposal:

You have the right to possess data about you. Regardless of what entity collects the data, the data belong to you, and you can access the data at any time. Data collectors thus play a role akin to a bank, managing the data on behalf of their "customers."

You have the right to full control over the use of your data. The terms of use must be opt-in and clearly explained in plain language. If you are not happy with the way a company uses your data, you can remove those data, just as you would close your account with a bank that is not providing satisfactory service.

You have the right to dispose of or distribute your data. You have the option to have data about you destroyed or redeployed.

At the World Economic Forum over the past five years, I have helped curate a discussion of these basic tenets among politicians, CEOs of multinational corporations, and public advocacy groups in the U.S., the European Union and around the world.

As a result, regulations in the U.S., the E.U. and elsewhere (such as the new U.S. Consumer Privacy Bill of Rights) are already giving individuals greater control over their data while also encouraging increased transparency and insight in both the public and private spheres.

LIVING LABS

FOR THE FIRST TIME in history, we can see enough about ourselves to build social systems that work better than the ones we have always had. Big data promises to lead to a transition on par with the invention of writing or the Internet.

Of course, moving to a data-driven society will be a challenge. In a world of unlimited data, even the scientific method as we typically use it no longer works: there are so many potential connections that our standard statistical tools often generate nonsense results. The standard scientific approach gives us good results when the hypothesis is clear and the data are designed to answer the question. But in the messy complexity of large-scale social systems, there are often thousands of reasonable hypotheses; it is impossible to tune the data to all of them at once. So in this new era, we will need to manage our society in a new way. We have to begin testing connections in the real world far earlier and more frequently than we ever have before. We need to construct "living labs" in which we can test our ideas for building data-driven societies.

One example of a living lab is the open-data city we just launched in Trento, Italy, with cooperation from the city government, Telecom Italia, Telefónica, the research university Fondazione Bruno Kessler and the Institute for Data Driven Design. The goal of this project is to promote greater idea flow within Trento. Software tools such as our openPDS (Personal Data Store) system, which implements the New Deal on Data, makes it safe for individuals to share personal data (such as health details or facts about their children) by controlling where their information goes and what is done with it. For example, one openPDS application encourages the sharing of best practices among families with young children. How do other families spend their money? How much do they get out and socialize? Which preschools or doctors do people stay with for the longest time? Once the individual gives permission, such data can be collected, anonymized and shared with other young families via openPDS safely and automatically.

We believe that experiments like the one we are carrying out in Trento will show that the potential rewards of a data-driven society are worth the effort—and the risk. Imagine: we could predict and mitigate financial crashes, detect and prevent infectious disease, use our natural resources wisely and encourage creativity to flourish. This fantasy could quickly become a reality—our reality, if we navigate the pitfalls carefully. 〽

MORE TO EXPLORE

Society's Nervous System: Building Effective Government, Energy, and Public Health Systems. A. Pentland in *Computer,* Vol. 45, No. 1, pages 31–38; January 2012.

Personal Data: The Emergence of a New Asset Class. World Economic Forum, January 2012. www.weforum.org/reports/personal-data-emergence-new-asset-class

The New Science of Building Great Teams. Alex "Sandy" Pentland in *Harvard Business Review;* April 2012.