
PART I

Introduction to the Pulse: A New Kind of Instrument

COPYRIGHTED MATERIAL
<http://www.pbookshop.com>

<http://www.pbookshop.com>

CHAPTER 1

A New Era for Measuring and Predicting Society

The digital traces of you and everyone you know are transforming our understanding of human behavior.

—*New Scientist*, July 24, 2010

Something important is changing in how we as a society use computers to mine data. . . . As more diverse sensors become pervasive, wireless networking becomes more widespread, and new algorithms are developed, a global sensor network monitoring much of humanity might emerge.

—Tom Mitchel of Carnegie Mellon as quoted in, *Science*, December 2009

The first half of the twenty-first century is seeing the emergence of a new type of scientific instrument that will measure the biggest and most important trends in society. More people spend more time doing more things online and as they do so, they leave behind a vast digital record. By using the combined “digital footprints” of the entire world, sociologists, psychologists, economists, and even physicians are learning to measure aspects of our society that would otherwise be virtually invisible. The analysis of these digital traces will generate new discoveries in every area of research that uses it. The volume of the data will detect new patterns in markets, public opinions, and even health and safety—and the information will be available fast enough to be actionable.

Publicly available data about the frequency of certain searches on Google have been used to track flu outbreaks, consumer confidence, or unemployment levels faster than the government authorities are able to do so. The analysis of millions of “Tweets” on the micro-blogging site Twitter have been used to predict the popularity of a movie, presidential approval ratings and even stock market moves. Classified ads on sites like Craigslist

correlate with home foreclosures and unemployment. Data gathered passively from mobile phones can be used to track traffic patterns, illness, and even damage after an earthquake. And this is just the beginning.

Even reserved scientists describe this phenomenon as a revolution in the study of humanity and in the forecasting of humanity's large-scale behavior. These tools will fundamentally change how we track threats and opportunities. Ultimately, this new source of data will influence how some of the most important decisions are made by individuals, businesses and governments.

Only relatively recently have researchers been discovering how to track economic trends, flu outbreaks, and public opinions by using publicly available data left behind by a couple of billion Internet and mobile phone users. These digital footprints are left by the people of the world as they use the Internet and mobile phones to blog, buy things, search for information, connect with other people, and even as they play online. This vast data set is the first opportunity for many of the social sciences to work with a quantity of detailed statistics that rivals or even exceeds the data sets of, say, particle physics or astronomy. Just as weather forecasting becomes more accurate with increased numbers of sensors and improved computational power, forecasting the "weather" of society will become just as scientific.

The weather map analogy works at many levels. The real value of weather maps is not in seeing where the weather patterns are right now; it is in seeing where the weather *will be*. And the more macroscopic the map, the better able we will be to see the trends. Likewise, many of the threats that we have faced in the first decades after 2000—terrorism, financial chaos, epidemics, and more—could be better seen in advance if we had a kind of macro-level weather map for society. And many opportunities will be apparent only if we can see where the big trends are leading business or public opinion.

Individual observations about where someone lives or the output of one small business say almost nothing about big trends. You can't see the size and shape of a storm by examining a few raindrops; similarly, you can't see big changes in the economy or society by looking only at your own neighbors or coworkers. However, a vast number of these data points can show us patterns on a large scale.

The Internet has been used for great business opportunities, as a repository of all human knowledge, as the means of collaboration across the globe, and as the ultimate tool of public dialogue. *However, a few initial examples from academia notwithstanding—the Internet itself is almost entirely underutilized as a measurement instrument of society.* The primary method of tracking big trends in society is still the survey and, in the case of business, the system of financial reports. Business and governments understand the significance of using these traditional surveys to inform critical strategies.

By 2002, the U.S. Government alone was spending over \$4 billion per year on surveys to measure the economy and other aspects of society. The commercial sector was spending about \$15 billion per year on the same.¹

Traditional survey methods involve carefully designed studies that allow for straightforward statistical analysis. Unfortunately, they are expensive, slow, and—even with the rigorous methodologies—often don't capture what is really happening. They are plagued with small response rates, and in some cases what people say on surveys has little to do with reality. These methods often take weeks to months for the report to be finished—many times this is long after the data would have been useful. A striking example of the lack of timeliness of existing analytical methods was when the National Bureau of Economic Research (NBER) reported in September 2010 that the Great Recession had ended 15 months prior. Surely there is room for improvement.

The Internet is already many orders of magnitude larger than all the data collected by governments and businesses using traditional surveys. The social networking site Facebook alone processes 25 terabytes of data each day. This is about 1,000 times the amount of data delivered by the U.S. mail daily. The largest survey in the world—the US Census—produces less data than this *in a decade*. In 2010, the microblogging service Twitter gave the Library of Congress 167 terabytes of data representing over 6 billion “tweets.” (A tweet is a single string of text, 140 characters or less, posted by one of the over 30 million Twitter users. Each one is a tiny daily diary entry saying something about what that person is thinking and doing.) And these sources still only comprise part of the publically-visible digital data we leave behind.

The data of the Internet is, of course, unstructured and in some ways more difficult to analyze than a purpose-built survey. Fortunately, new methods for analyzing this data scientifically are evolving even as the volume of data continues to grow at a fantastic rate. The collective digital record of our behavior has now grown to a point where it can reflect something useful about the system's users. Now we can see the equivalent of the weather map for public opinion, the economy, our health, and our anxieties.

This Book's Web Presence

This book, in keeping with the spirit of connecting to the Pulse, also has an online presence at www.pulsethenewscience.com. Throughout the book, I defer to the Web site for more elaborate examples of analysis of the Pulse. There, the reader can download spreadsheets and see other examples of the Pulse, including links to recent examples.

An Emerging Science: What Is This New Thing?

This new availability of terabytes of social data and the emerging social science dedicated to analyzing it is too new to even have a widely accepted name. I've considered terms like "social data analytics" or "cybersociology" as names for this science. I've thought about using "digital buzz" or "net pulse" to describe the trends that are made visible with the analysis. Still, these don't seem to fully capture the phenomenon and its magnitude.

Perhaps science fiction got close to the idea. In his 1951 science fiction classic, *Foundation* (Gnome Press Publishers), Isaac Asimov imagined "psychohistory"—a futuristic branch of mathematics applied to history and sociology. He described how it could be used to compute probabilities of macroscopic behaviors in society similar to how a physicist can predict the overall behavior of a gas without predicting the behavior of individual atoms. Psychohistory seems related to the emerging science this book discusses, but Asimov was vague on the source of data used to feed the formulas of psychohistory. What makes a real science possible is not just the math, but some way to collect the data, preferably a lot of it.

One intriguing term I've heard to refer specifically to the recent availability of the data is "social data revolution" (SDR). This phrase was coined by Andreas Weigend at Stanford University and it seems to encompass the essence of both the new science and the new social phenomenon. His term conveys both the suddenness and magnitude of the availability of social data from sources like Amazon book reviews, Twitter, eBay auctions and Facebook.

His credentials in business, science, and academics give Weigend a unique perspective on this topic. His PhD is in physics and his early work as an undergraduate was at CERN where he analyzed the data coming from particle detectors. This background in the unique challenges of analyzing large amounts of data are what led him to his position as the Chief Scientist of Amazon and later to his position as director of the Social Data Lab at Stanford.

He told me that SDR was first proposed not as a name for some grand new era but simply as the name for a course he was teaching at Stanford. He believed the previous name of the course, "Data Mining and E-Commerce" was too narrow and bland to cover the content the course was evolving into. He explained that his proposed name was accepted as both more comprehensive and inspiring: "In 2009, the department decided 'Social Data Revolution' captures its essence better and renamed the class. It was interesting how the new title attracted more creative, entrepreneurial students."

Another term I like, as you can tell from the title of the book, is simply the capitalized *Pulse*, since this is like the heartbeat of the macro-level

One Small Example of the Pulse

For a quick, concrete example of how to view one tiny aspect of the Pulse online, try this:

1. Go to www.google.com/trends to find the Google Trends site.
2. In the search field, type in the word “coupons.”
3. You will see a chart that shows how frequently this term was searched on Google ever since 2004.
4. Note that a sudden increase in searches on this word occurred in the Christmas season in 2007 followed by a bigger spike a year later—at the beginning of the financial crisis in 2008.

This is a simple and useful example of the Pulse because the frequency of searches on this term seems to correlate with certain economic conditions. (People use more coupons when their finances are squeezed.) In Chapter 5, we spend more time on how to interpret and use this data as well as account for other possible explanations for the sudden increase in search activity on this term. Later chapters investigate many other ways to use the Pulse.

behavior of the world. I credit Alan Mislove of Northeastern University in Boston with the first use of the world Pulse in this context (more on him in Chapter 7). Unlike “digital pulse” or “Internet buzz,” the term *Pulse* deemphasizes the specific technology and connotes only a single, massive rhythm. However, I may use these terms or perhaps “macro-pulse” to mean about the same things as Weigend’s SDR and Mislove’s Pulse. I’ll define the Pulse and its various synonyms as: *the collective, macroscopic trends which can be scientifically inferred by harnessing publicly accessible data from the Internet.*

Here are three important characteristics of this concept that are both explicit and implied from this definition:

1. *Practical.* In addition to being a powerful tool for academic research, this emerging science has practical applications. The Pulse is computed using sound mathematical methods, but it uses real-world data to solve real-world problems.
2. *Public.* The Internet is the data source. The Pulse is based on all data we leave on the Internet in some form that is publicly accessible, including

data that was initially gathered directly or indirectly by user input or passively gathered by other sensors employed by end users (e.g., the Global Positioning System [GPS] on mobile phones).

3. *Big.* We are focusing on aggregation of the data to spot big, macroscopic patterns and trends, not the identification of individuals or their private data.

If tapping into and using the Pulse seems a little abstract, you can take a quick look at one small example of this Pulse right now. (See “One Small Example of the Pulse.”) There are many more ways to access the Pulse than what is shown in this example. Besides analyzing searches on Google, we can analyze auctions on eBay, comments on Twitter, and much more. This is just one concrete example to start with.

What The Pulse Isn't

I explicitly define each of the three attributes as part of the Pulse in order to help distinguish this field from concepts that may be confused with it. The first attribute, practicality, helps to separate this field from some purely theoretical areas of social science that attempt to use mathematical models to describe social interactions. Many theories, even very quantitative ones, already exist for describing the large-scale behavior of groups of individuals. Yet, a theory without data is only half of science.

“Mathematical sociology” has been around since the 1960s. In the 1990s, researchers started using massive computational methods to simulate systems of human interaction. These areas of study became known as computational social science (CSS) or computational sociology. CSS might be part of the academic lineage of the Pulse, but it is not the whole story. Until very recently, CSS had so little data and so much theory it was more of a branch of mathematics than an empirical science. Even key researchers in CSS recognize this fact. (More on this later.)

The requirement of practicality means that the Pulse must support real-world decisions with real-world data. What makes the Pulse an empirical science as opposed to purely a mathematical curiosity is the new availability of data from the Internet. What makes it part of an applied science is how we focus on practical implications for businesses and governments.

Keeping practicality, public availability, and macroscopic scale squarely in mind, we also contrast the Pulse with a few other seemingly related but entirely different concepts. The areas that might be the easiest to confuse with this topic are issues related to the four areas listed next. The other related-sounding topics have already been addressed in plenty of other published material, none of which actually addresses harnessing the Pulse.

What the Pulse Is *Not* (Necessarily or Entirely) Related To

1. Diminishing privacy
2. Using the Internet for lead generation
3. The use of business tools commonly referred to as business intelligence and predictive analytics
4. Online versions of traditional survey methods

The topic of diminishing privacy is one that I frequently find to be associated with the topic of this book. As I speak on this subject at conferences around the United States, I encounter professionals who think that what we are really talking about is the growing intrusion on privacy by tracking our digital footprints. The concern is that exploiting the Pulse necessarily means diminishing privacy and, therefore, we shouldn't use the Pulse for *ethical* reasons. There are two reasons why this particular concern is not justified.

First, almost all of the data needed for this kind of analysis is highly aggregated; individual behaviors are lost. The Google Trends example from "One Small Example of the Pulse" shows nothing of the behavior of individuals. Google may know who is searching for what, but as long as it keeps that information confidential, the rest of us will not be able to see specific searches by other individuals. Whether you or I ever searched on a given term is lost in the data. Yet this highly aggregated data is all we need since we are focusing on the "macro," not the "micro"—the storm front, not the raindrops.

The second issue regarding the privacy concern is that where the data is detailed to the level of particular individuals—such as, for example, a review you write about a book on Amazon using your real name—it is explicitly voluntary. The expectation of privacy does not exist when we choose to use our real names on Twitter, Facebook, Amazon reviews, blogs, and so forth (especially when we include photos and home addresses). The fact is that more and more people are volunteering more information about themselves for public consumption than ever before. You may not be willing to divulge so much private information, but that doesn't mean we shouldn't learn what we can from those who are willing. Visualizing big trends with the Pulse is a way to make good use out of what would otherwise be little more than a cycle of exhibitionism and voyeurism.

At least some of the public is even willing to share their locations through "location-sharing" apps on mobile phones. Again, you may find this is too revealing for your taste, but many thousands—perhaps millions—will volunteer this information, and it can be used for a greater good. Individuals may prefer to reveal their location data for personal convenience (e.g., to have your phone tell you about nearby restaurants you might like or about sales at retail stores). Even so, this data can be shared in a way that doesn't

actually reveal individual locations. Like the raindrops in a storm, even this data can be blended together to make the radar map of the storm front. Both laws and technology will change over time in a way that may modify how we can collect data, but aggregation and deliberate volunteering of data will ensure that there is plenty of data to analyze regardless of privacy trends.

Privacy is a different issue from the Pulse, but it is still an important issue—and not just because of threats like identity theft. There are legitimate concerns that the use of vast quantities of social data could be exploited by oppressive regimes. If you are a dictator fearful of a popular resistance, you certainly want access to the Pulse. Evgeny Morozov, author of *The Net Delusion: The Dark Side of Internet Freedom*, argues that anyone who believes that the Internet is only a tool of democracy to be feared by authoritarian rulers is a naïve “digital utopian.”² He uses China and Iran as examples of governments who have effectively used the Internet to track the actions of citizens and to spread propaganda. The mining of social data has already been used against the citizens of a country.

On the other hand, authoritarian regimes sometimes fear social networking technology. Civil revolts in Iran in 2009 and in Tunisia, Egypt, Yemen, and Algeria in 2011 were facilitated by social networks. Egypt’s government cut off all Internet and mobile phone service in an attempt to control the organizing power these tools had for protestors. Perhaps with analysis of the Pulse, Egyptian authorities could have *anticipated* the unrest and kept a dictator in power. Like anything else, the Pulse can be used for good or evil. What the reader needs to prepare for is that, either way, the Pulse is happening.

Another area that needs differentiation from the Pulse is various “lead generation” methods and the isolation of individual tweets as news. The Pulse certainly can be used for strategic marketing decisions, but this book doesn’t focus on issues like identifying individual leads with data from the Internet. There already is a lot of literature about how to exploit the Internet for lead generation, so there is no need to replicate that body of work.

In a similar vein, many companies already “farm” their online product reviews to try to address proactively negative reactions through direct outreach to disaffected individuals. News organizations also cull blogs and Twitter for public reactions to events and, in some cases, to get wind of new stories. However, in all of these cases, there is a lack of pattern-seeking other than purely anecdotal observations. Again, such an approach focuses on the raindrops instead of the storm front.

Now, if CNN discontinued the segment where it randomly picks tweets that it thinks convey a sentiment about a topic and instead showed a CNN “sentiment index” (SI), then CNN is truly using the Pulse. CNN would

report the sentiment index just like it reports the Dow Jones Industrial Average, and it would be validated by being used to predict other measurable phenomenon, such as the results of a traditionally gathered Gallup poll. Journalists could report that the CNN SI on some aspect of the economy is rising slower or faster than expected. Since the SI would be validated by comparison to historical economic trends, it could be an indicator of things to come. Now, that would be much closer to real news.

Exploiting viral marketing methods, however, does seem like a topic worthy of the social data revolution. A “viral” video or Internet myth grows rapidly in a way similar to viruses in the microbiology sense. A person “catches” it from another person, who then passes it on to many others, who each pass it on to many others. Particularly entertaining items, for example, make people want to share them. Viral videos, conspiracy theories, jokes, and news items travel throughout the cyberworld in a way that has patterns—patterns worth investigation for both academic and practical purposes.

Now let’s distinguish the Pulse from the analysis of confidential corporate or government databases. The data of interest to us is what could be accessed by the public or could be gathered by anyone—and I’m not talking about security breaches as in the case of Wikileaks in 2010. Confidential data known only to an individual and a company that gathers the information—such as the phone company that has records of your phone calls, or the bank that records financial transactions you specifically make—is very useful to those firms but not to everyone else. It is possible that those firms could use that data to make their own forecasts of the economy or other social trends. Still, since that is not the kind of data most other individuals or firms would have access to, it will be of limited value to decision makers outside of firms that own those databases and it is not considered part of the Pulse.

The public accessibility feature is also a reason why we need to contrast the Pulse with the well-known business tools of data mining, business intelligence (BI), and predictive analytics (PA). Data mining certainly is one of the tools used in the Pulse with the caveat that, in this context, the data mining is entirely outward-looking—searching on the user-generated data available on the Internet—not a review of internal corporate or government files. BI and PA capture part of the spirit of this new science. They are all about analyzing huge amounts of corporate data to try to inform management about the state of the business and possible trends. However, so far, these terms have been used mostly for proprietary and internal databases. Software that has been developed for these purposes eventually will need to combine analysis with the Pulse, and some developers of BI and PA tools have already begun to do so.

Finally, we are not focusing so much on explicit data gathering, such as taking online surveys. There are tools available that are effectively online versions of traditional survey methods. Although the method of administering the survey is, no doubt, much more streamlined than paper-based methods, online surveys are never going to have the same volume of data that Twitter, Google, Facebook, or eBay gather every day. When people are using those sites, they are leaving behind useful information that could be used for BI—but users don't go to those sites just because they want to give some analyst information about their lives.

The Major Areas of the Macro-Pulse

In Part Two, this book examines the world of the digital Pulse by looking at the new tools, methods, and recent research in six broad and overlapping categories of activities. This taxonomy divides the Pulse into areas according to how people leave behind their digital footprints. You can remember the categories in this little rhyme:

Surf, Friend, Say, Go, Buy, and Play.

- *What we surf.* Tools like Google Trends and Google Adwords can be used to determine what the public is searching for. When there is an increase in searches including the words “unemployment” or “coupon,” we also see that economic conditions are worsening. How frequently people in one city search on “flu symptoms” has been shown to be a good indicator of where flu outbreaks will occur. There are also tools that tell us about the popularity of specific Web sites. This too tells us something about what people need, want, or fear.
- *Whom we friend.* “Friend” has become a verb (aka “befriend”) in the world of online social networking, and we can use the information you share about who you friend to gather other insights into the structure of society. Using the details of social networks, we can learn about the propagation of gossip and social fads as well as more serious issues like the adoption of new technologies and the spread of diseases. The kinds of social networks people keep tell us something about who they are and what they do. In some cases, the information can even be used to determine criminal threats.
- *What we say.* When we leave our 140-character-long comments on Twitter, comment on a blog, write a customer review for a book, or post a comment on a friend's wall on Facebook, we are leaving behind an important anthropological record of society. What we say in these continuously updated media have been found to agree

closely with Gallup poll research on public opinions of politicians or consumer confidence—but with the added benefit of being nearly real time.

- *Where we go.* A large number of people have already begun to voluntarily transmit their locations from their GPS-equipped mobile phones. Some early research showed that volunteers providing GPS data from mobile phones had entirely different social behaviors from those they had reported on surveys. (This research calls into question many of the findings of traditional survey methods.) Even at a highly aggregated level that reveals nothing about individual locations, this data can be very useful. Where the crowds go may say something about their economic activities and even their opinions and moods.
- *What we buy.* The auctions on a site like eBay are available for anyone to see. By a measure of total items sold, eBay is one of the world's largest retailers, and—unlike most major retailers—it posts every sale it makes the moment the sale is made. The total number of transactions on eBay exceeds the transactions on the New York Stock Exchange and Nasdaq combined. Yet, unlike those financial indexes, eBay presents information about direct consumer purchases. Likewise, Amazon gives us a glimpse of the interests of people when it provides the sales rank of a book (how well that book sells compared to other books). If the Amazon sales ranks of books about, say, buying a new home show that a large number of books on that topic are selling better, this may tell us something about not just what people are doing but what they *intend* to do.
- *How we play.* Perhaps the most underinvestigated area of the macro-Pulse is the collaborative and competitive behavior on sites where we play together. Research shows that our moods do affect the choices we make, even in the games we play. A game session can be considered a rich data stream of individual decisions. Given that some games have millions of players, it is an anthropological record too great to ignore. In addition, there are relatively simple game applications (apps) for Facebook and iPhones. Some of these games have proven to be particularly popular (e.g., Farmville for Facebook). There is even the prospect of creating games *especially* for data collection for the Pulse.

This taxonomy is not perfect, and these areas are not necessarily mutually exclusive or collectively exhaustive. By their nature, many topics of interest could be part of more than one category shown here. Gaming, for example, covers not just what we play but what we buy and say on the Internet. Still, I think this is a good starting point for a very quickly evolving topic. Now in the following paragraphs, we will briefly explore the implications of the Pulse for decision makers.

What Does All of This Mean for Us?

Even though the research about how big trends can be seen by mining the data left behind by the world's Internet use is exciting, the purpose of this book will ultimately be about practical consequences. Businesses and governments can use this emerging science to their benefit—and not just to find better ways to promote products and ideas. Yes, that is included in the Pulse, but the social data revolution will have even more significant effects on society as a whole.

A good analogy for this might seem to be the impact e-business has had on business—but that may not be a close analogy at all. The effects of measuring the Pulse may be closer to how the adoption of common units of measure, laws, and language affected trade. The existence of these standards not only facilitated more trade and better trade but made new kinds of trade possible. In a similar vein, flying at night and sharing the air with thousands of other aircraft going different directions would not merely be less efficient without real-time data from instruments and radar; it wouldn't be *possible* at all. Microscopes didn't just make it easier to see small things; they made possible microbiology and, ultimately, most of modern medicine.

The Pulse will be similarly impactful. Businesses and governments won't just make faster and more accurate decisions by tracking big-picture trends in real time with these tools. They will do things completely differently. Information from the Pulse will become part of models that are dynamically simulating and forecasting businesses and the consequences of management decisions. Perhaps the furthest reaching effects of the Pulse will be beyond what we can imagine at this point. Even so, it will have at least the following four types of impact:

1. *Decisions based on responses to macro-trends will be faster.* We won't have to wait weeks—and certainly not 15 months as in the NBER report on the end of the recession—for indications of changes in fundamental economic factors, health conditions, and public opinions.
2. *In some cases, the Pulse will be more accurate than traditional methods of collecting data about major trends.* While the traditional methods like polls will continue to be used to calibrate and validate the Pulse, the Pulse will avoid some problems that plague traditional surveys.
3. *Trends that otherwise would not have been seen at all will be visible.* Traditional surveys have to be purpose-built. In other words, we have to have an idea of what we are looking for in advance, and then we have to collect that specific data. The Pulse offers a way to see trends that no one even knew to look for when the data was generated.

4. *Basic models of society will change.* Our ability to investigate and respond to the environment more quickly and accurately has implications for organizational structure, logistics, finance, and virtually every other part of business, government, and the study of humanity. This may be the greatest impact of the Pulse.

What economic benefits will the Pulse eventually help us realize? Our only benchmarks for this kind of fundamental shift are things like mass production, powered transportation, electricity, telecommunications, and computing. These innovations combined to revolutionize agriculture, health, life spans, per capita incomes, and even human rights. Our economy has grown by more than a factor of 100 since the start of these changes to society. The benefits of the Pulse are wide open.

Now, in the next three chapters, let's see how we got here and what "here" looks like. Chapter 2 puts tracking the big picture in context by showing a history—both recent and ancient—about how civilization has tried to ascertain the big picture and how that knowledge changed things. We review the history of how the trends would be approximated with traditional surveys and how early information technology had an impact.

Chapter 3 explains the birth of the digital Pulse—how several disciplines converged with a vast and rapidly expanding body of data from Internet users. We review not just how the number of connected people is growing but how much more they do in the Internet, how much more diverse the connected population is becoming, and how influential the "connected" are in society. We outline some of the major developments in the Pulse.

Chapter 4, the last chapter of Part I, will describe some of the relevant structure of the Internet as it applies to the Pulse. The Pulse relies on a system of incentives for individuals providing data to services and services making some of this data public. The basic methods used by developers to mine data and how this data is distributed around the Internet are discussed.

Notes

1. University of Michigan Ann Arbor Survey Research Center, as stated in a 2002 National Science Foundation grant proposal, "Identifying Causal Mechanisms Underlying Nonignorable Unit Nonresponse through Refusals to Surveys."
2. E. Morozov. *The Net Delusion: The Dark Side of Internet Freedom*, PublicAffairs, 2011

<http://www.pbookshop.com>