

The Uses of Big Data in Cities

Luis M. A. Bettencourt

SFI WORKING PAPER: 2013-09-029

SFI Working Papers contain accounts of scientific work of the author(s) and do not necessarily represent the views of the Santa Fe Institute. We accept papers intended for publication in peer-reviewed journals or proceedings volumes, but not papers that have already appeared in print. Except for papers by our external faculty, papers must be based on work done at SFI, inspired by an invited visit to or collaboration at SFI, or funded by an SFI grant.

©NOTICE: This working paper is included by permission of the contributing author(s) as a means to ensure timely distribution of the scholarly and technical work on a non-commercial basis. Copyright and all rights therein are maintained by the author(s). It is understood that all persons copying this information will adhere to the terms and constraints invoked by each author's copyright. These works may be reposted only with the explicit permission of the copyright holder.

www.santafe.edu



SANTA FE INSTITUTE

The Uses of Big Data in Cities

Luís M. A. Bettencourt
Santa Fe Institute
1399 Hyde Park Rd
Santa Fe, NM 87501, USA.

Email: bettencourt@santafe.edu
@BettencourtLuis

September 17, 2013

Abstract:

There is much enthusiasm currently about the possibilities created by new and more extensive sources of data to better understand and manage cities. Here, I explore how big data can be useful in urban planning by formalizing the planning process as a general computational problem. I show that, under general conditions, new sources of data coordinated with urban policy can be applied following fundamental principles of engineering to achieve new solutions to important age-old urban problems. I also show, that comprehensive urban planning is computationally intractable (i.e. practically impossible) in large cities, regardless of the amounts of data available. This dilemma between the need for planning and coordination and its impossibility in detail is resolved by the recognition that cities are first and foremost self-organizing social networks embedded in space and enabled by urban infrastructure and services. As such the primary role of big data in cities is to facilitate information flows and mechanisms of learning and coordination by heterogeneous individuals. However, processes of self-organization in cities, as well as of service improvement and expansion, must rely on general principles that enforce necessary conditions for cities to operate and evolve. Such ideas are the core of a developing scientific theory of cities, which is itself enabled by the growing availability of quantitative data on thousands of cities worldwide, across different geographies and levels of development. These three uses of data and information technologies in cities constitute then the necessary pillars for more successful urban policy and management that encourages, and does not stifle, the fundamental role of cities as engines of development and innovation in human societies.

Keywords: Urban Planning & Policy, Engineering, Feedback Control Theory, Computational Complexity, Self-Organization, Science of Cities.

1. New Opportunities for the Use of Big Data in Cities

How does one measure a city? By the buildings that fill its skyline? By the efficiency of its rapid transit? Or, perhaps, by what Jane Jacobs called the “sidewalk ballet” of a busy street? Certainly these are the memorable hallmarks of any modern city or metropolitan area. But a city’s true measure goes beyond human-made structures and lies deeper than daily routine. Rather, cities and metro areas are defined by the quality of the ideas they generate, the innovations they spur, and the opportunities they create for the people living within and outside the city limits. - Judith Rodin, 2013 [1].

The rise of information and communication technologies (ICT) [2,3] and the spread of urbanization [4] are arguable the two most important global trends at play across the world today. Both are unprecedented in their scope and magnitude in history, and both will likely change the way we live irreversibly. If current trends continue, we may reasonably expect that the vast majority of people everywhere in the world will live in urban environments within just a few decades and that information technologies will be part of their daily lives, embedded in their dwellings, communications, transportation and other urban services.

It is therefore an obvious opportunity to use such technologies to better understand urbanism as a way of life [5,6], and to improve and attempt to resolve the many challenges that urban development entails in both developed and developing cities [4,7].

Despite its general appeal, the fundamental opportunities and challenges of using big data in cities have, in my opinion, not been sufficiently formalized. In particular, the necessary conditions for the general strategic application of big data in cities need to be spelled out and their limitations must also be, as much as possible, anticipated and clarified. To address these questions in light of the current interdisciplinary knowledge of cities is the main objective of this perspective¹.

Before I start, it is important to emphasize that the use of quantitative data to better understand urban problems and to guide their solutions is not at all new. In the United States – and in New York City in particular – we have been building detailed statistical pictures of urban issues for over a century. Jacob Riis’ influential “How the other half lives” [11] derived much of its persuasive power from the use of statistics, e.g. by giving numbers of deaths in tenements in New York City to the person. The NYC-RAND Institute in the 1970s [12] used detailed urban statistics, modeling and computation developed for wartime and corporate management to determine resource allocation, especially for New York City’s Fire Department [13]. Thus, today’s *Smart Cities*

¹ In this paper I do not discuss issues of ethics or privacy related to big data, nor of the political or corporate potential pitfalls of using big data to manage cities. This is not because these issues are not important but because here I wish to focus on the fundamental possibilities to manage cities based on data, without such complications. It can probably be said that any of these considerations will work to limit the scope and effectiveness of the use of data to address urban issues and will, as a consequence, make the case for its use less compelling. Several authors have recently written eloquently about several of these issues [8,9,10].

movement [14-28] needs to be placed in perspective: Are the achievements made possible by modern data and information technologies fundamentally different from what was possible in the past? Or are they the extensions of old ideas and procedures, albeit with greater scope and precision?

To answer these questions I formalize the use of data in urban policy and management in light of the conceptual frameworks of engineering [29]. I show, in the next section, that this allows us to identify the necessary conditions for the effective use of data and policy to address a large array of urban issues. In this way I will demonstrate that, in principle at least, modern information and communication technologies are opening entirely new windows of opportunity for the application of engineering solutions to cities.

But the problems of cities are always primarily about people. Social and economic issues in cities define what planners call ‘wicked problems’ [30], a term that has gained currency in policy analysis well beyond the urban context. These are the kind of issues that are not expected to yield to engineering solutions, for specific reasons [30] that break the assumptions of feedback control theory [29]. In section 3 I revisit the issue of ‘wicked problem’ in light of computational complexity theory [31] to make the formal argument that detailed urban planning is computationally intractable. This means that solutions requiring the knowledge and forecast of chains of detailed behaviors in cities have the fundamental property that they become *practically impossible* in large cities, regardless of the size of data available. This clarifies the central dilemma of urban planning and policy: that planning is clearly necessary to address long-term issues that span the city, e.g. in terms of services and infrastructure, and yet that the effects of such plans are impossible to evaluate *a priori* in detail.

In Section 4 I resolve this dilemma. The key is the nature of self-organization of social and economic life in cities and the development of a general quantitative understanding of how such processes operate in cities as vast networks, spanning large spatial and temporal scales. Theory recognizes that most individual details are irrelevant to describe complex systems as a whole, while identifying crucial general dynamics [32]. For example, a city exists and functions even as people change their place of residence, jobs and friendships. The development of this kind of urban theory too follows from increasing data availability in many cities around the world, both in terms of observations and, wherever possible, from experiments.

I then discuss how the increasing integration of limited-scope engineering solutions with the dynamics of social self-organization in cities, articulated by the long and large views afforded by urban theory provide the appropriate context to understand and manage cities, while allowing them to play their primary function in human societies of continuing to evolve socially and economically in open-ended ways.

2. The Nature of Big Data Solutions in Cities

It is a profoundly erroneous truism [...] that we should cultivate the habit of thinking what we are doing. The precise opposite is the case. Civilization advances by extending the number of important operations, which we can perform without thinking about them.

- Alfred Whitehead, 1911 [33].

Relatively simple-minded solutions, enabled by precise measurements and prompt responses, can sometimes operate wonders even in seemingly very complex systems where traditional policies or technologies have failed in the past.

An example, not immediately related to managing cities, illustrates this point: self-driving cars. I have recently attended a workshop at the Kavli Institute for Theoretical Physics in Santa Barbara, CA, as part of a group of interdisciplinary researchers discussing issues of artificial intelligence, algorithms and their relation to neuroscience. In a particularly memorable lecture, Prof. Richard Murray from Caltech² explained how his team engineered a self-driving vehicle that completed DARPA's *Grand Challenge* in 2005 [34] and was now deploying it on the road in cities³. Richard described the array of diverse sensors for determining surrounding objects and their speed, for geo-locating the vehicle, recognizing the road and its boundaries and planning its path. I interjected asking whether the vehicle learned from its experience and, to my surprise, the response was no: it was all "hardwired" by the team at Caltech, using principles of feedback control theory. I protested that surely the hardest problem faced by a self-driving car was not planning its path, recognizing the road or measuring surrounding traffic – but all the stupid things that people do on the road: "Doesn't the car need a model of human behavior? That is surely the hard part!"

The answer, surprisingly, is no, not at all! The reason is that while people surely do reckless things behind the wheel, they do so, from the car's perspective, at a truly glacial pace: While humans behave unpredictably on the time scale of seconds, the car's electronics and actuators respond in milliseconds or faster. In practice, this allows the car more than enough time to take very basic 'hardwired' evasive actions, such as breaking or getting out of the way. In fact, presently self-driving cars have a safety record that exceeds that of humans [35,36].

The lesson from this story is that relatively simple solutions that require no great intelligence can, under specific conditions, solve very hard problems. The key is fast and precise enough measurement and adequate simple reactions. In other words, sometimes you don't need to be very smart if you are fast enough.

² This paragraph is based on my personal recollection of Richard Murray's lecture. Any inaccuracies or embellishments are mine not the speaker's.

³ Self-driving cars or autonomous vehicles (most famously developed by Sebastian Thrun's team from Stanford U., now at *Google*) are no longer news in urban environments. *Google* deploys a dozen cars on a typical day in the Bay Area [35].

This is the logic of modern engineering, and more precisely of feedback control theory [29]. If we know the desired operating point for a system (e.g. *avoid collision*) and we have the means to operate on the system, as we observe its state change, via feedback loops, we can under general conditions turn it into a simple problem. The crucial conditions are to be able to measure and recognize potential problems just as they start to arise (car approaching) and act to make the necessary corrections (break or get out of the way, thereby *avoiding collision*). The crucial issue is that of temporal scales; every system has intrinsic timescales at which problems develop, a few seconds in the case of a car on the road. Cycles of measurement and reaction that avoid such complex problems by simple means must act well within this window of opportunity.

We now see more clearly why a new generation of bigger data may offer radically novel solutions to difficult problems. Modern electronics are now so fast in comparison to most physical, biological and social phenomena that myriads of important policy problems are falling within this window of opportunity. In such circumstances, models of system response can be very simple and crude (they can typically be linearized) [29]. Thus, the engineering approach conveniently bypasses the complexities that always arise in these systems at longer temporal or larger spatial scales, such as the need to develop models of human behavior. In this way a difficult and important problem can be solved essentially *without theory* [37]. This is the (potential) miracle of big data in cities.

Most examples of urban management and policy in cities that use data successfully have this flavor, regardless of being implemented by people and organizations or by computer algorithms. Table 1 presents a summary of important urban issues, where I attempted to roughly characterize their typical temporal and spatial scales and the nature of their operating points, or outcomes.

For example, think of urban transportation systems, such as a bus network, in these terms [38]. Buses should carry passengers who wait a few minutes to be transported over a few kilometers. Measuring the time in between buses at each stop, possibly together with the number of passengers waiting, gives the planner the basis for a feedback control solution: Communicate with buses to enforce desired standards of service, quickly place more or fewer units in service where these parameters start to deviate from the ideal metrics, and the quality of service as measured by per person waiting times will improve. This type of strategy can be operated intuitively by human dispatchers but possibly can also be implemented by an ICT algorithm [38] with access to the necessary measurements and actions. Feedback control theory provides the framework for the development and optimization of any of these solutions.

Similar procedures could be devised for water or power supply management and for their integration with public transportation systems [39,40]. Traffic management, trash collection, infrastructure maintenance (building, roads, pipes, etc) could also generally be thought of, and integrated together, in analogous ways.

| Problem | Timescale | Spatial Scale | Outcome Metric |
|--|------------------|----------------------|-----------------------|
| Transportation (buses, subway) | minutes | meters | simple |
| Fire | minutes | meters | simple |
| Epidemics (HIV, Influenza) | years, days | city-wide | simple |
| Chronic Diseases | decades | city-wide | simple |
| Sanitation | years | city-wide | simple |
| Crime | minutes | meters | simple |
| Infrastructure (roads, pipes, cables) | days | meters | simple |
| Traffic | minutes | Meters-km | simple |
| Trash collection | days | meters | simple |
| Education | decades | city-wide | complex |
| Economic Development | decades | city-wide | complex |
| Employment | years | city-wide | complex |
| Poverty | decades | neighborhood | complex |
| Energy and Sustainability | years | city-wide | complex |
| Public housing | years-decades | neighborhood | complex |

Table 1. Urban issues, their temporal and spatial scales and the character of their associated metrics. Outcome metrics are to be interpreted on the time scale in the table. On longer timescales socioeconomic issues, characterized by long times, become part of every problem.

And there is no reason in principle why certain stubbornly difficult social problems, such as crime, may not be responsive to patrolling and other law enforcement strategies that follow similar principles. Crime is local and its outcomes clear enough. Exciting new experiments in several cities, notably in Los Angeles [41,42], suggest that much.

Recent insights by innovative corporations [43-46] and policy strategists [47] have correctly anticipated that flows of information underpin such coordinated solutions. Thus, progress in these problems is fundamentally an ICT problem, enabled by simple actions or policies that correct the states of systems towards optimal performance.

But other urban issues, especially those that are primarily social or economic acquire, as seen through this lens, a different character because their operating points are not well defined or because their dynamics are rather diffuse and play out over large temporal or spatial scales and are the result of large chains of social events, Table 1. Thus, it has remained difficult to create engineering solutions to problems of education, public housing⁴, economic development, poverty or sustainability at the city level.

⁴ Notably, public housing policy in places such as Hong Kong [48] or Singapore [49] has fared uncharacteristically well when compared to western cities. These successes are likely due to a combination of an adaptable design, where apartments can be expanded as households become

Public health issues often sit at an intermediate level of complexity. Analogously to crime, contagious diseases are characterized by simple metrics and by local processes of social contact between individuals. Thus, containment or eradication of cholera or of contagious diseases for which vaccination exists, represent some of the greatest successes of urban policy⁵. But conditions that play out over longer times⁶ and may have more complex and diffuse social causation, such as chronic diseases or even HIV/AIDS have proven thus far more difficult.

Thus, the simplicity of performance metrics expressed as objective quantitative quantities and knowledge of their proximate causes in space and time are the crucial conditions for successful engineering inspired solutions. The important point to always bear in mind, however, is that these quantities are relative to the properties of the controlling system – the policy maker or the algorithm – such as their response times, so that with technological and scientific evolution we may hope to fulfill Alfred Whitehead’s maxim quoted at the beginning of this section, of achieving progress through the increasing automation of solutions to once intractable problems⁷.

3. The Planner’s Problem

In order to describe a wicked-problem in sufficient detail, one has to develop an exhaustive inventory of all conceivable solutions ahead of time. The reason is that every question asking for additional information depends upon the understanding of the problem-- and its resolution --at that time. [...] Therefore, in order to anticipate all questions (in order to anticipate all information required for resolution ahead of time), knowledge of all conceivable solutions is required.
- Rittel and Webber, 1973 [30].

The scenarios developed in the previous section paint a well-defined path for the use of data in cities, at odds with many decades of urban theory and practice. While many physical and infrastructural aspects of cities appear at first sight manageable through engineering practices, many social and economic issues are in fact ‘wicked problems’

more prosperous, careful matching between housing characteristics and socioeconomic household capacities and close monitoring of housing problems and prompt response, in terms of conflict resolution, maintenance and repair.

⁵ Dealing with the problem of cholera, by John Snow in London in the 1840s-50s, constitutes one of the first examples of a statistical approach to public health, though one thoroughly based on “shoe-leather epidemiology” [50].

⁶ An exception to this argument is influenza. It is a ‘fast’ disease that has defied public health response times. Its symptoms are generic and it typically is a large-scale problem, affecting 20-30% of a population over large spatial regions.

⁷ The arrow of progress made possible by engineering and automation does not always run in the same direction. The self-driving car solution that motivated this section works only as long as most cars are driven by people. Once most cars are self-driven and share the same fast response times, new instabilities will develop. A preview of such problems can be seen in stock markets today, where it has been estimated that maybe over 80% of trades are made by algorithms on ultrafast time-scales [51]. Often such algorithms act on news written by other ultrafast algorithms.

[30]. Over the long run the social and infrastructural aspects of the city are always entangled.

Wicked problems were originally defined by two urban planners, Rittel and Webber, in 1973 [30]. The essential character of wicked problems is that they cannot be solved in practice by a (central) planner. I shall reformulate some of their arguments in modern form in what I like to call the “Planner’s Problem”. The *Planner’s Problem* has two distinct facets: i) the knowledge problem and ii) the calculation problem.

The *knowledge* problem refers to the information (ultimately as data) that a planner would need to map and understand the current state of the system; the city in our case. While still implausible, it is not impossible to conceive information and communication technologies that would give a planner, sitting in a ‘situation room’, access to detailed information about every aspect of the infrastructure, services and social lives in a city. Privacy concerns aside, it is conceivable that the lives and physical infrastructure of a large city could be adequately sensed in several million places at fine temporal rates, producing large but potentially manageable rates of information flow by current technology standards⁸. In this way the knowledge problem is not a showstopper.

The *calculation* problem, however, alluded to in the quote at the beginning of this section, refers to the computational complexity to perform the actual task of planning, in terms of the number of steps necessary to evaluate all possible scenarios and choose the best possible course of action. This problem is analogous to playing a complex game, like chess, but with millions of pieces, each following different rules of interaction with others, against millions of opponents, on a vast board. It will therefore not surprise the reader that the exhaustive approach of evaluating each possible scenario in a city is impractical as it involves the consideration of impossibly large spaces of possibilities.

I formalize this statement in the form of a theorem, and provide mathematical details and the sketch of a proof in the Appendix. The result of considering the combinatorial possibilities of a city is the creation and selection of the best detailed urban plan would require a computation involving $O[P(N)]$ steps in a city with population size N , where

$$P(N) \simeq P_0 e^{N^{1+\delta} \ln N},$$

where $\delta \simeq \frac{1}{6}$ and P_0 is a constant, independent of N . For a city of a million people this is

$P(N) = P_0 10^{6 \times 10^6}$, a truly astronomical number much, much larger than all the atoms in the Universe. Thus, the Planner’s Problem is practically impossible.

⁸ For the largest city in the world, Tokyo with 35 million inhabitants, we could imagine sensing each individual and a corresponding number of places, including different infrastructure and services using, say, 100 million data streams, sampled at one byte each once a second (a faster rate may be desirable for some applications, slower in others). This results in information flows of the order of 10^9 bits per second (Gbps) that are already handled by Gigabit Internet and current generation computing. So while still daunting these numbers can already be handled using current technologies.

I want to emphasize that the sketch of a proof given in the Appendix relies on the formulation of the Planner's Problem as the choice of the best spatial and social configuration for the city that maximizes asset of urban metrics. A more limited conceptualization of the problem will likely have lower computational complexity, but may not guarantee finding the best plan. In such a case, planning must rely on heuristics and cannot guarantee absolute optimality.

Given these caveats, I have shown that planning the city in detail becomes computationally impossible in all but the smallest towns. This indicates that the use of complex models [39,53] in the detailed planning and management of cities has its limits and cannot be exhaustively mapped and solved in general, regardless of how much data a planner may acquire. How then should we think of planning under these limitations, and what becomes the role of bigger and better data in cities?

4. Self-organization, Information and the Role of a Science of Cities

Provided that some groups on earth continue either muddling or revolutionizing themselves into periods of economic development, we can be absolutely sure of a few things about future cities. The cities will not be smaller, simpler or more specialized as cities of today. Rather, they will be more intricate, comprehensive, diversified and larger than today's and will have even more complicated jumbles of old and new things as ours do. - Jane Jacobs, 1970 [54]

Part of the answer to the Planner's Problem comes from economics, where the social (central) planner faces a similar, if more limited, predicament of organizing economic markets. The detailed information necessary to run a city in terms of planning the lives of people and organizations is, of course, best known to those agents themselves. Even more importantly, the motivation and capacity to make good decisions and act on this information resides with the very same agents. Thus, cities are self-organizing and depend crucially not on the detailed instructions from the planner's 'situation room' but rather on the integration and coordination of myriads of individual decisions, made possible by suitable information flows⁹.

In this sense, cities have always been the primary creators and users of information and communication technologies, from the daily newspaper, to postal mail, to the telegraph or the (cellular) telephone [57]. It should then come as no surprise that new ICTs will primarily act to enable cities, rather than compete with them. In fact, the problem that ICTs address in specific niches, is the general problem solved by cities [58]: That is, the integration and coordination of people and organizations through information sharing in

⁹ The original argument for self-organization in the context of economics was made e.g. by Friedrich Hayek in the essay "The Use of Knowledge in Society" [55], a title that the present manuscript echoes. The central argument is that economies are primarily about information and that the price system of economics is one (good) way of achieving the necessary self-organization in terms of the allocations of goods and services, labor and capital. This point is also emphasized by Paul Krugman in his 1992 Ohlin lectures [56].

vast social networks that can confer benefits to individual agents and to society as a whole.

This then brings us to another perspective on the use of big data in cities, radically different from the engineering framework of the previous sections. A vital function of information and communication technologies is in enabling new and better ways for people to be social. In this respect big data and associated technologies play only a supportive role to social dynamics, not a prescriptive one. Data and technologies then do not create or solve urban problems so much as they enable people and social organization to address them better. This role of ICT has been discussed in the literature on smart cities [19,21,25,26]. Especially important is the potentially more rapid access to civic and economic participation of poor or excluded populations, particularly in developing cities and nations [59].

The second part of the answer stems from the fact that bottom-up self-organization is clearly not enough for a well functioning city. Many developing cities, think of Mumbai [60], have self-organization in spades. Similarly, a flash-mob is as much the product of self-organization as a new art form or a new business. For cities to work, general properties and constraints that define the city on larger scales must be at play. Such features, common to all cities, are the focus of urban theory.

Scientific theory is not about the details of a particular practical problem; it is about general principles that apply across many problem instances. Given how different cities are from each other and how much they change over time, to what extent can a scientific theory of cities be possible or useful?

To illustrate this point consider another great triumph of science and engineering: the Moon shot. It is not completely inconceivable that the Apollo 11 mission could have been pulled off in the absence of knowledge of the laws of mechanics and gravity. One could imagine controlling rocket thrusters using feedback by judging whether the spacecraft was leaving Earth, approaching the Moon at the right speed not to crash, etc. But of course this is not at all what happened: such solutions would be very difficult in terms of engineering and computation at the time, especially because of the fast time scales and large energies necessary for lift-off. Instead, we launch rockets based on their general properties as prescribed by physical theory, the mass of the Earth and of the spacecraft, the rotation of the planet, the speed necessary at lift-off, etc. This set of calculations gives a general strategy for launching *any* spacecraft from Earth and to place them in desired orbital positions.

What follows this general solution is a detailed, but much more limited, use of engineering principles and self-organization to fine-tune solutions to the details of the problem, such as the conditions for Moon landing given the particular spacecraft's speed at that moment, its distance to the surface, the human operator past actions, etc.

Thus, theory can enormously reduce the size of an engineering problem and in many cases makes solutions possible. In this way, urban theory sets the general parameters for

any city to work as a whole as a bottom-up self-organizing process mediated by social networks with certain general properties and enabled by infrastructure and services that follow general performance metrics [58]. The details of these processes must then be measured and managed, as appropriate, in every situation.

The present state of urban theory is fast evolving and is starting to approach the status of an interdisciplinary science. New, bigger data that allows for the comparative quantitative analysis of thousands of cities worldwide has been the key in determining the general properties of cities and their variation due to both general conditions of development and local histories [58,61,62]. We now have a quantitative framework that connects the general structure of cities as vast social networks, to their general properties of infrastructure networks and land use [58]. In this way we are uncovering what kind of complex system cities really are [6,58,63] and the general conditions in terms of infrastructure, energy and resource use, etc, that allow cities to function optimally as distinct open-ended social systems, capable of socioeconomic growth and innovation in human societies¹⁰.

5. Discussion

The last great technological advancement that reshaped cities was the automobile (some might argue it was the elevator). In both cases, these technologies reshaped the physical aspects of living in cities – how far a person could travel or how high a building could climb. But the fundamentals of how cities worked remained the same. What’s different about the information age that has been ushered in by personal computers, mobile phones and the Internet is its ability to reshape the social organization of cities and empower everyday citizens with the knowledge and tools to actively participate in the policy, planning and management of cities.

- Christian Madera, 2010 [66]

Developed cities today are social and technical complex systems characterized by historically unprecedented levels of diversity and temporal and functional integration. This growing individual specialization and interdependence makes large cities extremely diverse and crucially relies on fine temporal and spatial integration and on faster and more reliable information flows. These processes lie at the core of what makes cities the economic and cultural engines of all human societies.

A city of 35 million people, like present day Tokyo, is made possible by extremely efficient transportation, reliable energy and water supply and good social services, including a very low level of conflict. Many developing urban areas in the world today will have to replicate and improve upon these metrics if they are to fulfill their promise to become world cities and enable the pace of human development expected by millions of their inhabitants.

¹⁰ While recent progress in urban theory has allowed us to better understand cities as multi-level interacting networks and as processes of self-organization across scales, more work needs to be done regarding heterogeneous aspects of the city (inequality of incomes, variations across neighborhoods [64]) and the evolution of cities over time, including the processes of social development and economic growth [54,65].

I have showed here that the increasing separation of time-scales between information and communication technologies and the natural pace of human behavior will enable many fundamentally new engineered solutions that may solve important age-old urban problems. Thus, new sensing, communications and computational technologies will play a vital role in enabling new urban scale efficiencies and city socioeconomic growth. In this sense we should expect that with urban development, services and infrastructure should become increasingly engineered, invisible, predictable and automatic. How resilience to infrastructure and service failures may be maintained in these circumstances is an important open problem. Important as they are such solutions should always be understood to be short-term and to have a limited scope, as they must continue to adapt and enable socioeconomic development over the long run.

The world's most vibrant and attractive cities are not usually the same places where the buses run impeccably on time. While improvements in infrastructure and urban services are absolutely necessary for cities to function better, they are not the fundamental sources of social development or economic growth. These, more fundamental functions of cities, rely on processes of social self-organization and on the fulfillment of general conditions that allow cities to operate effectively as multi-level open-ended evolving networks. It is, curiously, this more fundamental character of cities, at play across scales from the individual to the city as a whole and from seconds to centuries, that creates and solves the *Planner's Problem*.

In this light, the key to making cities even smarter, through new uses of information and communication technologies, is not very different from what it has always been. Big data creates more opportunities, not less, for cities to accentuate their urban character: The primary uses of big data in cities must then be to continue to enable the creation of new knowledge by more people, not replace it.

Acknowledgements

I thank the participants of the workshop "How Far Can "Big Data Take Us Towards Understanding Cities?" at the Santa Fe Institute September 19-21, 2013, where a version of this paper was presented. This research was partially supported by the Rockefeller Foundation, the James S. McDonnell Foundation (grant no. 220020195), the John Templeton Foundation (grant no. 15705), the Army Research Office (grant no. W911NF1210097), the Bill and Melinda Gates Foundation (grant OPP1076282) and by a gift from the Bryan J. and June B. Zwan Foundation.

References

1. J. Rodin in the foreword to B. Katz, J. Bradley, *The Metropolitan Revolution: How Cities and Metros Are Fixing Our Broken Politics and Fragile Economy*. (Brookings Institution Press, Washington DC, 2013).

2. M. Hilbert, P. López, The World's Technological Capacity to Store, Communicate, and Compute Information, *Science* 332: 60–65 (2011).
3. V. Mayer-Schonberger, K. Cukier, *Big Data: A Revolution That Will Transform How We Live, Work, and Think* (Eamon Dolan/Houghton Mifflin Harcourt, New York NY, 2013)
4. UN-Habitat, *Planning Sustainable Cities – Global Report on Human Settlements 2009*. Available <http://www.unhabitat.org/content.asp?typeid=19&catid=555&cid=5607>
5. L. Wirth, Urbanism as a Way of Life, *The American Journal of Sociology*, 44: 1-24 (1938).
6. L.M.A. Bettencourt, The Kind of Problem a City is, in D. Offenhuber and C. Ratti. (eds) *Die Stadt Entschlüsseln: Wie Echtzeitdaten Den Urbanismus Verändern: Wie Echtzeitdaten den Urbanismus verändern*. (Birkhauser, Heidelberg, 2013). Available (in english) <http://www.santafe.edu/media/workingpapers/13-03-008.pdf>
7. UN-HABITAT, *The Challenge of Slums: Global Report on Human Settlements* (2003).
8. J. Polonetsky, O. Tene, Privacy and Big Data: Making Ends Meet, 66 *Stanford Law Review* 2. Available <http://www.stanfordlawreview.org/online/privacy-and-big-data/privacy-and-big-data>
9. d. boyd, K. Crawford, Critical Questions for Big Data, *Information, Communication and Society* 15: 662-679 (2012).
10. McKensey Global Institute, Big data: The next frontier for innovation, competition, and productivity (2011). Available http://www.mckinsey.com/insights/business_technology/big_data_the_next_frontier_for_innovation
11. J. Riis, *How the Other Half Lives: Studies among the Tenements of New York* (Kessinger Publishing, 2004). Originally published in 1890. Available <http://www.authentichistory.com/1898-1913/2-progressivism/2-riis/index.html>
12. P. L. Szanton, Analysis and Urban Government: Experience of the New York City-Rand Institute, *Policy Sciences* 3: 153-161 (1972).
13. J. Flood, *The Fires: How a Computer Formula, Big Ideas, and the Best of Intentions Burned Down New York City and Determined the Future of Cities* (Riverhead Books, New York, NY, 2011)

14. D.V. Gibson, G. Kozmetsky, R. W. Smilor (eds.), *The Technopolis Phenomenon: Smart Cities, Fast Systems, Global Networks*. (Rowman & Littlefield, New York NY, 1992).
15. M. Castells, *The Informational City: Information Technology, Economic Restructuring, and the Urban-Regional Process* (Blackwell, Oxford, UK, 1989).
16. M. Batty, The Computable City. In: Fourth International Conference on Computers in Urban Planning and Urban Management, Melbourne, Australia, July 11-14 (1995). Available http://www.acturban.org/biennial/doc_planners/computable_city.htm
17. S. Graham (ed.), *The Cybercities Reader*. (Routledge, London, 2004).
18. W. J. Mitchell, *Me++: the cyborg self and the networked city*, (The MIT Press, Cambridge MA, 2004)
19. H. Partridge, Developing a human perspective to the digital divide in the Smart City (2004). Available <http://eprints.qut.edu.au/1299/1/partridge.h.2.paper.pdf>
20. R. Giffinger, et al., Smart Cities Ranking of European Medium-Sized Cities. Centre of Regional Science, Vienna University of Technology, Vienna, Austria (2007). Available <http://www.smart-cities.eu/>
21. J. M. Shapiro, Smart cities: quality of life, productivity, and the growth effects of human capital, *The Review of Economics and Statistics* 88: 324-335 (2008).
22. S. H. Lee, J. H. Han, Y.T. Leem, T. Yigitcanlar, Towards ubiquitous city: concept, planning, and experiences in the Republic of Korea. In: T. Yigitcanlar, K. Velibeyoglu, S. Baum (eds.) *Knowledge-Based Urban Development: Planning and Applications in the Information Era*, pp. 148–170. (Information Science Reference, Hershey, 2008).
23. A. Caragliu, C. Del Bo, P. Nijkamp, Smart Cities in Europe. VU University Amsterdam, Faculty of Economics, Business Administration and Econometrics (Research Memoranda 0048, Amsterdam, The Netherlands, 2009).
24. C. Harrison, B. Eckman, R. Hamilton, P. Hartswick, J. Kalagnanam, J. Paraszczak, P. Williams, Foundations for Smarter Cities, *IBM J. Res. Develop.* 54: 1-16 (2010).
25. N. Taewoo, T. Pardo, Conceptualizing Smart City with Dimensions of Technology, People, and Institutions, *The Proceedings of the 12th Annual International Conference on Digital Government Research* (2011). Available <http://dl.acm.org/citation.cfm?id=2037602>

26. M. Batty, K.W. Axhausen, F. Giannotti, A. Pozdnoukhov, A. Bazzani, M. Wachowicz, G. Ouzounis, and Y. Portugali, Smart cities of the future, *The European Physical Journal Special Topics* 214: 481–518. (2012).
27. S. Roche, N. Nabian, K. Kloeckl, C. Ratti, Are ‘Smart Cities’ Smart Enough? In: *Global Geospatial Conference*. Global Spatial Data Infrastructure Association (2012). Available <http://www.gsdi.org/gsdiconf/gsdi13/papers/182.pdf>
28. R. Papa, C. Gargiulo, A. Galderisi, Towards an Urban Planner’s Perspective on Smart City, *TeMA* 1: 5-17 (2013).
29. K. J. Åström, R. M. Murray, *Feedback Systems: An Introduction for Scientists and Engineers* (Princeton University Press, Princeton NJ, 2008). Available http://www.cds.caltech.edu/~murray/amwiki/index.php/Main_Page
30. H.W.J. Rittel, M. M. Webber, Dilemmas in a General Theory of Planning, *Policy Sciences* 4: 155-169 (1973).
31. C. More and S. Mertens, *The Nature of Computation* (Oxford University Press, New York NY, 2011)
32. P. W. Anderson, More is Different, *Science* 177: 393-396 (1972).
33. Alfred N. Whitehead *An Introduction to Mathematics* (Williams & Norgate, London, 1911).
34. Team Caltech at the DARPA *Grand Challenges*: http://archive.darpa.mil/grandchallenge/TechPapers/Cal_Tech.pdf
35. Google driverless car: http://en.wikipedia.org/wiki/Google_driverless_car
36. M. M. Moore, B. Lu, Autonomous Vehicles for Personal Transport: A Technology Assessment (June 2, 2011). Available <http://ssrn.com/abstract=1865047> or <http://dx.doi.org/10.2139/ssrn.1865047>
37. C. Anderson, The End of Theory: The Data Deluge Makes the Scientific Method Obsolete, *Wired* 06.23.08 http://www.wired.com/science/discoveries/magazine/16-07/pb_theory
38. J.S.J. Ren , W. Wang , S. S. Liao, Optimal Control Theory in Intelligent Transportation Systems Research - A Review, arXiv:1304.3778 [cs.SY]

39. C. Harrison, *Roads to Smarter Cities* (2013). Available upon request from the author.
40. MASDAR City. <http://www.masdarcity.ae/en/>
41. M.B. Short, M.R. D'Orsogna, P.J. Brantingham, and G.E. Tita, Measuring and modeling repeat and near-repeat burglary effects, *Journal Quantitative Criminology* 25: 325-339 (2009).
42. J. Rubin, Stopping crime before it starts, *Los Angeles Times* August 21, 2010. Available
<http://articles.latimes.com/2010/aug/21/local/la-me-predictcrime-20100427-1>
43. IBM Global Business Services, *A Vision of Smarter Cities: How Cities Can Lead the Way into a Prosperous and Sustainable Future*, (IBM Institute for Business Value, Somers, NY, 2009)
44. Cisco's smart city framework (2012):
<http://www.cisco.com/web/about/ac79/docs/ps/motm/Smart-City-Framework.pdf>
45. Siemens' Smart City in Vienna: Model Project Aspern (2013).
http://www.siemens.com/innovation/en/news/2013/e_inno_1319_1.htm
46. Smart Cities Council: <http://smartcitiescouncil.com>
47. European Union's Report of the Meeting of Advisory group
ICT Infrastructure for energy-efficient buildings and neighborhoods for carbon-neutral cities (2011)
http://ec.europa.eu/information_society/activities/sustainable_growth/docs/smart-cities/smart-cities-adv-group_report.pdf
48. D. J. Dwyer (Ed.), *Asian Urbanization, a Hong Kong Casebook*. (Hong Kong University Press, Hong Kong, 1971).
49. W. Fernandez, *Our Homes: 50 Years Of Housing A Nation* (Housing and Development Board, Singapore, 2011).
50. S. B. Johnson, *The Ghost Map: The Story of London's Most Terrifying Epidemic - and How it Changed Science, Cities and the Modern World* (Riverhead Books, New York NY, 2007)
51. N. Johnson, G. Zhao, E. Hunsader, H. Qi, N. Johnson, J. Meng, B. Tivnan, Abrupt rise of new machine ecology beyond human response time, *Nature Scientific Reports* 3, Article number: 2627 (2013).
52. Historical Internet Traffic Rates: http://en.wikipedia.org/wiki/Internet_traffic

- 53 Y. Portugali, *Complexity, Cognition and the City* (Springer, Heidelberg, 2011)
54. J. Jacobs, *The economy of cities* (Vintage, New York, NY, 1970).
55. F. Hayek, The Use of Knowledge in Society, *American Economic Review*. 35: 519-530 (1945).
56. P. Krugman, *Development, Geography, and Economic Theory (Ohlin Lectures)* (MIT Press, Cambridge MA, 1997).
57. P. Hall, *Cities in Civilization* (Pantheon Books, New York NY, 1998)
58. L. M. A. Bettencourt, The Origin of Scaling in Cities, *Science* 340: 1438-1441 (2013).
59. C. Andris, L.M.A. Bettencourt, Development, Information and Social Connectivity in Côte d'Ivoire, Santa Fe institute Working Paper Paper #: 13-06-023. Available <http://www.santafe.edu/media/workingpapers/13-06-023.pdf>
60. S. Mehta, *Maximum City: Bombay Lost and Found* (Vintage, New York NY, 2005).
61. L. M. A Bettencourt, J. Lobo, D. Helbing, C. Kühnert, G. B. West, Growth, Innovation, Scaling, and the Pace of Life in Cities. *Proc. Natl. Acad. Sci. U.S.A.* 104, 7301-7306 (2007).
62. L.M.A. Bettencourt, J. Lobo, D. Strumsky, G.B. West, Urban Scaling and Its Deviations: Revealing the Structure of Wealth, Innovation and Crime across Cities. *PLoS ONE* 5(11): e13541 (2010).
- 63 J. Jacobs, *The Death and Life of Great American Cities* (Random House, New York, 1961).
64. R. J. Sampson, J. D. Morenoff, T. Gannon-Rowley, Assessing “Neighborhood Effects”: Social Processes and New Directions in Research, *Annual Review Sociology* 28:443–478 (2002).
65. E. L. Glaeser, H. D. Kallal, J. A. Scheinkman and A. Shleifer, Growth in Cities, *Journal of Political Economy* 100: 1126-1152 (1992).
66. C. Madera “The Future of Cities in the Internet Era” , *Next City* (2010). Available <http://nextcity.org/daily/entry/the-future-of-cities-in-the-internet-era>
67. J. M. Harris, J. L. Hirst, M. J. Mossinghoff, *Combinatorics and Graph Theory* (Springer, New York NY, 2008).

68. M. Schläpfer, et al. The Scaling of Human Interactions with City Size. Available <http://arxiv.org/abs/1210.5215>

69. K. J. Arrow, A Difficulty in the Concept of Social Welfare, *Journal of Political Economy* 58: 328–346 (1950).

Appendix

The Planner's Problem

Definitions:

Consider a city with N people defined as a connected social network [57]. This means that every person in the city can be connected to any other through a series of social links between third parties, however long. The social network that characterizes the city is G , which is made up of K social links. Both the general structure of G and its size K are functions of N and of other urban factors such as the costs (e.g. transportation, crime, time) and benefits (wages, information) of social interaction in that city.

The city's "performance" is monitored in terms of a set of quantities Y_1, Y_2, \dots, Y_M . These urban metrics may vary from case to case but I simply require that they are measurable quantitatively. The important property of each Y_i is that it is a function of the city's social configuration, that is $Y_i(G)$. The metrics Y_i can measure familiar things, such as the size of city's economy (GDP), the state of its public health, urban crime rates, the cost of transportation, indices of environmental sustainability, etc, including more subjective quantities such as happiness, or satisfaction. I show below that what is measured and how many indicators are considered is not substantially important. The defining property is that they are functions of the social networks that make up the city, G .

Problem Statement:

Given these definitions, the *Planner's Problem* is to choose the plan associated with the best G , subject on K remaining constant, so that a suitable function of the city's metrics Y_i is maximized.

Sketch of proof:

The Planner's Problem is computationally intractable:

Without more specific knowledge, consider a city of N people characterized by all possibilities for social connections between them, K_p . For example, in the simplest case that these are undirected links we have $K_p = N(N-1)/2$. Within this space the planner

needs to determine the plan that assigns K connections optimally. The number of plans, P , therefore is given by the number of combinations of K in K_p [67] or

$$P(K_p, K) = \frac{K_p!}{K!(K_p - K)!} = e^{\ln K_p! - \ln K! - \ln(K_p - K)!} \equiv e^{B(K_p, K)}. \quad (\text{S1})$$

For large K_p , K I can use Stirling's approximation to the factorial

$$\ln n! = n(\ln n - 1) + O(\ln n), \quad (\text{S2})$$

in the exponent $B(K_p, K)$. After some algebra I obtain,

$$B(K_p, K) \simeq K \left[\ln \left(\frac{K_p}{K} \right) - 1 \right] \simeq K \ln \left(\frac{K_p}{K} \right). \quad (\text{S3})$$

Now, urban scaling theory [58] and cell phone data [68], tell us that $K = K_0 N^{1+\delta}$, where K_0 is a constant of order a few connections, and $\delta \simeq \frac{1}{6}$. The number of potential connections may vary somewhat depending on additional constraints on the problem, but the simplest assumption, given above, is that $K_p \sim N^2$ (Metcalf's Law).

Thus, with these choices I obtain,

$$B(N) \simeq B_0 N^{1+\delta} \ln N. \quad (\text{S4})$$

Then, the number of plans that need to be evaluated to make the best choice is

$$P(N) \simeq P_0 e^{N^{1+\delta} \ln N}, \quad (\text{S5})$$

which grows *faster than exponentially* with the number of people in the city. This is an astronomical number, for example, for a million people $N = 10^6$, I obtain

$$P(N) = P_0 e^{1.4 \times 10^7} = P_0 10^{6 \times 10^6}, \quad (\text{S6})$$

much, much larger than all the atoms in the universe ($\sim 10^{82}$).

The number of computational steps needed to evaluate the best plan is set by the evaluation of the quantities Y_i for each plan. Regardless of details, this can be done in $\sim M \times P$ steps, plus the sorting of these plans by these values. The fastest sorting algorithms require at least P steps ($P \log P$ is more typical), so I conclude that the complete algorithm would have a computational complexity proportional to at least

$P(N)$. Such algorithm would scale with the population size of the city N , faster than exponentially and would be impractical in all but the smallest towns.

Algorithms that scale exponentially or faster with the size of the input set are considered technically intractable, that is, they cannot be evaluated in practice as this set gets large. Thus, the (computational) planner's problem, under the present set of assumptions, is intractable.

Notes:

1. The detailed combinatorial arguments involved in a specific algorithm for the Planner's Problem are likely to be more constrained if we include considerations of time, budget, etc. This may result in effectively a smaller number of potential connections being viable, though it is uncertain whether such a judgment can be made a priori, without the consideration of all possibilities as done above. In any case note that the leading term in the exponent B is set by the number of social connection K in the city, not K_p . As such, these considerations play a sub-leading role in the calculation.
2. Given a number of urban indicators $M \geq 3$ there are well-known difficulties in producing an ordering of improvements that satisfies most inhabitants in the city. This is related to Arrow's Impossibility Theorem [69]. Such problems arise in addition to the computational complexity of the *Planner's Problem*.
3. Decisions and actions by individuals and organizations based on information they obtain in their urban environments effectively solve the *Planner's Problem*. They do so by "parallelizing" the problem and its inference and computations through the simultaneous pursuit of local adaptations that, in principle at least, can maximize each agent's preferences under constraints (budgets, knowledge, time, etc). This self-organizing dynamics is not guaranteed to produce the best outcomes city-wide, though. A more formal argument for self-organization in cities and what it can and cannot achieve remains an open problem that includes, but also needs to transcend, some of the foundational results in microeconomic theory.