Title
ICT4ICTD: Computational Social Science for Digital Development

Author
Hilbert, Martin

Publication Date
2015

DOI
10.1109/hicss.2015.258

Peer reviewed

ICT4ICTD: Computational Social Science for Digital Development

Martin Hilbert, University of California, Davis

Author's version of Hilbert, M. (2015). ICT4ICTD: Computational Social Science for Digital Development. In IEEE Computer Society (pp. 2145 – 2157). http://doi.org/10.1109/HICSS.2015.258

Abstract

While the ICT for Development (ICTD) community is well aware about the far-reaching changes introduced by the digital age, it is remarkably slowly getting used to the idea that digital tools also revolutionize its very own core business: research. Information and communication technology (ICT) is currently transforming the way knowledge is created and insights are obtained. This applies to inductive empirical inquiry (i.e. 'big data'), as well as deductive theoretical scholarship (i.e. 'agent-based computer simulations'). This article explores best practices of the application of such Computational Social Science in the field of development and contributes an informed perspective to intensify an outstanding discussion within the ICTD community. ICT should not only be used for development (ICTD), but also for the constant updating of our understanding of digital development, in order to fine-tune policies and project designs: 'ICT4ICTD'. On the one hand, the article shows that the consideration of this double role of ICT has the potential to significantly increase the impact of ICTD. On the other hand, developing countries and the ICTD community face important challenges when applying these tools, which should never be adopted uncritically.

1. Introduction

Digital Information and Communication Technologies (ICT) are currently revolutionizing the way research is carried out. This affects both main components of any scientific project: empirical work with data (the main driver of induction), and theoretical model building (the main driver of deduction).

Given that ever more of human conduct is taking place in digital networks, and given that digital conduct inevitably leaves a digital footprint, the social sciences currently have access to an unprecedented amount of data on the most diverse aspects of the social fabric and its development dynamics [1]. The catch-phrase here became big data [2-7], and its impact on the social sciences has been compared with the impact of the invention of the telescope for astronomy and the invention of the microscope for biology (providing an unprecedented level of detail about the system of interest). Confronted with such increase in the level of perceivable granularity in social dynamics, social scientists have an inevitable obligation to make use of it to inform analysis, policy and project design. Since ICT for development (ICTD) dynamics unavoidably involve ICT, most ICTD projects automatically produce such digital footprint.

While the opportunities of big data are enormous, especially for developing countries in which traditional statistics are scarce [5-7], they are subject to the same limitation as all statistics. Its ultimate limitation is known as the 'Lucas critique' in economics [8], as 'Goodhart's law' in finance [9] and as 'Campbell's law' in education [10]. All date back to 1976, when the Nobel Prize winning economist Robert Lucas criticized colleagues who used sophisticated statistics to make economic predictions ('econometrics') in order to inform policy making. He argued that no useful information can emerge from such analysis because "any change in policy will systematically alter the structure of econometric models" [8]. The argument is that all kinds of data (including econometric and 'big') are from the past (as 'real-time' as they might be), so any data analysis can only tell us about structures and dynamics from the past. When the past and the future follow the same logic, this is useful. However, if significant changes occur in the system's dynamic, empirical statistics are at best limited, if not deceiving. Development work has the explicit goal to create a future that is significantly different from the past. It aims at changing aspects of the *modus operandi* of the system. Considering the diversity of development settings, the outcome is very context dependent and almost always unique. This limits the usefulness of data from a specific case of the past.

In order to predict a future that has never been, theory-driven models are necessary. These allow variables to be adjusted with values that have never existed in statistically observable reality. ICT also acts as a game changer in this challenge. Computational simulations allow to set up theory-driven models that greatly expand the scope and level of sophistication of traditional 'paper-and-pen' models. While traditional models are only able to handle a very limited number of variables (at most a dozen or so), todays computational power allows creating mathematically formalized models with thousands and even millions of dynamic variables. Such computer simulations of artificial societies have no conceptual limitations on the achievable level of detail and accuracy. Most recent simulations are based on individual agents ('agent-based models'), resulting in an emergent interplay between bottom-up and top-down dynamics [11-13].

The combination of both is understood as Computational Social Science in this article [e.g. 1,14]. Sections 2 and 3 review the characteristics of these two aspects of Computational Social Science for development, while section 4 turns to the ensuing opportunities and challenges. A review of some 100 referenced articles inform the presented perspective.

2. Big data for development

The value unleashed by big data to inform decision has been referred to as "the new oil" [4] and recent literature has started to point to the important opportunities that big data opens up for development [5-7]. The OECD is convinced that "big data now represents a core economic asset that can create significant competitive advantage" [15] and even the UN Economic and Social Council has already reported to the UN Secretary General that "big data have the potential to produce more relevant and more timely statistics than traditional sources of official statistics, such as survey and administrative data sources" [16].

The crux of the big data paradigm is basically twofold. For one, it refers to new sources of data. The digital footprint created with each digital communication and transaction can replace traditional data sources (like surveys) with proxy indicators that correlate with the variable of interest. The benefit is the low cost and real-time availability of the digital proxy indicator. The epitome is Google's illustrious use of 50 million most common search terms as a proxy for the spread of the seasonal flu [17,18].

Secondly the notion of big data goes beyond the increasingly large amount of data itself, and focuses on methods of data analytics to inform intelligent decisions. Independent from the specific giga-, tera-, peta-, or exabyte scale, the big data paradigm argues to systematically place analytic treatment of data at the forefront of intelligent decision-making. The process can be seen as the natural next step in the evolution from the "Information Age" and "Information Societies" to "Knowledge Societies" [7]. Building on the digital infrastructure that led to vast increases in information, the big data paradigm focuses on converting this digital information into knowledge that informs intelligent decisions. Continuing with the previous example, Google processed an impressive 450 million different mathematical models in order to test for correlations between online search terms and flu outbreaks reported by official data. Eventually, 45 search terms were identified that outperformed traditional models of flu outbreak with real-time predictions [17].

2.1. Characteristics of big data

The big data paradigm can be characterized by some five general features [19]. Big data:

- replaces random sampling with the ambition to *capture all* there is (sampling n = universe N).
- is often accessible in *real-time*.
- is produced anyways as a low-cost and almost inevitable *digital footprint*.
- is messy and incomplete, which can be compensated by data redundancy from different sources, often called *data fusion*.
- uses exploratory data mining and *machine-learning models*, which replace the need for theory with plain pattern detection.

2.2. Big data and development

We will review concepts and applied examples of each of these five characteristics from the field of development.

2.2.1. Universal sampling. One the the biggest potential for big data in developing countries consists in mobile phone data. With a global penetration of over 95 % [20] and around 75 % access among those making US$1 per day or less [21], mobile phones became an important source of social data in developing countries. Their pervasiveness allows to access important features of society as a whole, without the need for sampling. The key enabler is the detection of the correlations between mobile phone usage and some other 'real-world' conduct [22]. On the one hand, understanding mobile phone usage provides insights into the nature of technology diffusion and the digital divide [23], as well as commercial telecom operators with "critical information for the personalization and adaptation of mobile-based services to the behavioral segments identified" [24; p. 36]. This might often even be the incentive for such studies. However, on the other hand, once this relation is understood, mobile phone records can also be used the other way around to infer demographic, socio-economic and other behavioral trades based on the continuously registered digital conduct.

For one, this has the potential to complement traditional statistical surveys. For example, the measurement of poverty levels is one of the politically most relevant statistic throughout the developing world. Using information from mobile phone call records through a plurality of base stations, it has been shown how the prediction of socioeconomic level in a geographic region can automatically be performed [24,25]. Prediction accuracy depends on the kind of variable (for example, predicting gender from mobile phone behavior is surprisingly tricky [26,27]), but generally it has been shown to be around 80-85% when using mobile call data records like call duration or frequency [24,26-28]. Such ideas can be fine-tuned for cases where more detailed digital footprints are available. For example, given that some 95 % of the mobile phones in developing countries are prepaid [21] and given that people put economic priority on recharging the phone, even under economic constraints [29], tracking the level of mobile phone recharging would provide a great source to measure the development of poverty levels in real time on a fine-grained geographic level [5].

Secondly, universal mobile phone sampling has also demonstrated its potential to obtain new insights. For example, mobile phone records from rural Kenya have been used to provide unprecedentedly detailed travel and migration patterns in low-income settings to understand the spread of malaria [30] and infectious diseases [31]; population movements following an earthquake and cholera outbreak in Haiti [32,33]; social responses to urban earthquakes in Mexico [34]; and charity and reciprocal aid among peers in Rwanda after the strike of a natural disaster [35]. Telecom companies already sell mobility profiles obtained from mobile phones to business clients, who can gain insights into consumer behavior in real-time [36].

2.2.2. Real-time shadow. Mobile phones are not only universal, but also provide real-time information. As such, the attenuation from radio signals when rain falls between cellular towers has been used as a big data source to measure the amount of rain that falls in an area [101]. Such real-time and large scale precipitation measurements can bring timely information to farmers, water resource managers, and climate researchers in the developing world where standard rain gauge networks are usually underdeveloped.

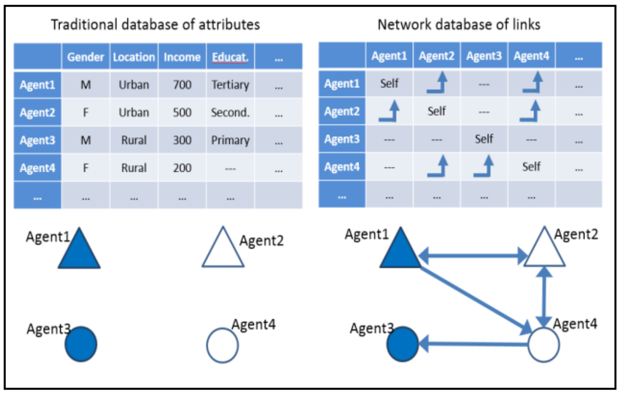
One of the most common real-time sources for big data is the incessant chatter in online social media. This source is especially important in developing countries, considering their acceptance and the wide arrange of content they provide. The top five leading countries in terms of Facebook users in 2013 included India, Brazil, Indonesia and Mexico [37], while in 2011 Kuwait and Brunei had more Twitter users per capita than the UK or U.S., Chile more than Canada, and Brazil more than France or Germany [38].

Twitter geolocated data has been used for the automatic identification of land uses [39], and the language content of Twitter microblogs has been used to approximate cultural identities, international migration and tourism mobility, including in countries like Malaysia, the Philippines, Venezuela and Indonesia [38]. Similarly, it has been shown that the 140 character long micro blogs from Twitter contained important information about the spread of the 2010 Haitian cholera outbreak up to two weeks earlier than official statistics [40]. Kalampokis et al. [41] investigated 52 articles that made use of social media sources to make social predictions. 13 of them made use of status updates from Facebook and Twitter to predict elections, 10 used web-search engine queries to make economic predictions, and 9 used input from blogs and review boards to predict the spread of disease and reactions to natural disasters.

2.2.3. Data as digital byproduct. The production of a digital footprint that serves as big data source is almost inevitable. As such, digital conduct has also shed light on previously under-investigated aspects of the social fabric. For example, it has provided visible and illustrative evidence the importance of social ties. This is not only useful to study social media, but extremely useful for attaining deeper insights into any kind of development dynamic, since development is as much about ‘who you are’, as ‘with whom you are’. The tools and techniques of social network analysis [42,43] allow to exploit the networked social fabric that development is made of. Concepts like ‘agent of change’, ‘intermediary’, ‘gatekeeper’, ‘broker’, ‘polarization’, ‘exclusion’, ‘marginalization’, ‘fragmentation’, ‘discrimination’, ‘social stability’ ‘social capital’, and ‘sphere of influence’ are mathematically precisely definable concepts in social network analysis and help to formalize the discussion about development.

Before the digital age it was costly and burdensome to obtain the required data and the adequate computational power to analyze the underlying networks (through computationally intensive matrix algebra). While traditional statistical analysis exclusively focuses on collecting and analyzing attributes of independent agents (e.g. income levels, demographics, location, and gadgets that belong to specific social agents), social network analysis additionally collects and analyses the network ties between these actors. This requires the creation of additional databases that describe social relations in matrix format (see Figure 1). Studying these linkages is complementary to traditional statistical methods. Unfortunately, official statistics and metrics produced to accompany development projects traditionally do not consider, nor analyze these second kinds of databases to inform decisions. Digital conduct often provides them gratis.

Figure 1: Schematization of traditional database (left) and network database (right)

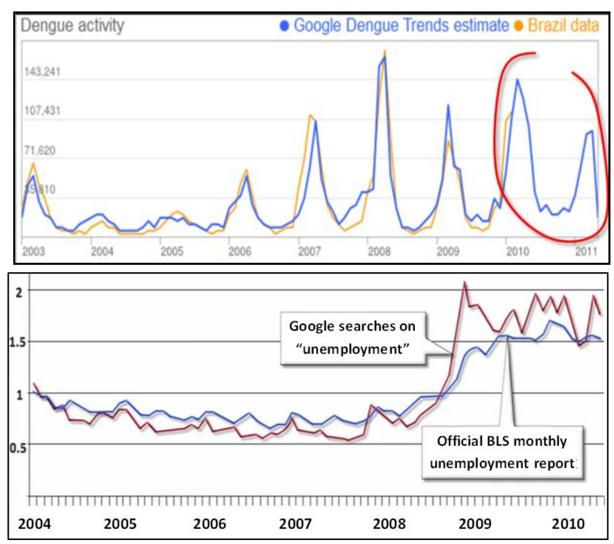


It has been shown that social network ties are important indicators of economic development on the micro- and macro-level. For example, there is a strong relation between economic development and the diversity of individuals' relationships [44] and of production networks of an economy [45]. Network approaches have already proven their usefulness to gain insights and guide policy in ICTD, such as in the design of media campaigns for reproductive health [46]; the general dynamic of the diffusion of innovations such as ICT [47,48]; the accelerated diffusion of innovations through pinpointed targeting of opinion leader [49], including empirical studies of the acceleration of the diffusion of microfinance [50]; inter-organizational cooperation and knowledge-sharing within the ICTD community [51]; and the role of telecenters and cybercafés in the creation of social capital in developing countries [52].

2.2.4. Data-fusion. Mobile, social media and network data (and others) can also been combined to fine-tune important indicators of development. For example, the current methods on measuring development are quite coarse-grained, which is exemplified by the most important development barometer, the United Nations Human Development Index [53]. While being a better reflection of development than any single variable, and surely highly innovative 25 years ago (even winning a Nobel Prize [54]), it merely feeds of four generic indicators: life-expectancy, adult literacy, school enrollment ratio, and Gross Domestic Product per capita. It has increasingly been subject of critique [55] and scholars have proposed and developed a large variety of alternatives, including so-called “happiness” indices, which are produced based on costly subject surveys [56].

Using big data, the Thomson Reuters MarketPsych Indices (TRMI) distills daily over 3 million news articles and 4 million social media sites through an extensively curated language framework [57]. It not only measures different emotions (such as optimism, confusion, urgency etc.), but also opinions (such price forecasts etc.) and specific topics (such as special events, etc.). The company provides 18,864 separate indices, across 119 countries, curated since 1998, and updated on a daily, or even minute basis. The result is a fine-grained, real-time assessment of the local, national or regional sentiment in terms of development relevant indicators such as wellbeing, happiness, content, and security, and even fear, stress, urgency, optimism, trust or anger, among others. The use of diverse sources results in a dataset that is messy and incomplete. In most big data exercises often not one single row of data is complete (not everybody provides social media feeds). However, data redundancy among different sources allows to make up for this fact by the complementary treatment of different sources.

Figure 2: (a) Google Brazil Dengue Activities [61]; **(b)** Google searches on unemployment vs. official government statistics from the Bureau of Labor Statistics [62].



2.2.5. End of theory. Most big data predictions are the result of data mining and machine learning techniques not informed by theory, but simply by pattern detection [58]. The amount of data allows such atheoretical predictions to work impressively well. Figure 2 illustrates this logic by showing how Google search word trend estimate on dengue trends in Brazil are able to identify dengue outbreaks while the official estimates from the Brazilian Ministry of Health are still missing (see Figure 2a, see also [59]). Figure 2b shows an even simpler model. It visualizes how closely the simple number of Google searches for the word “unemployment” in the U.S. correlates very closely with actual unemployment data from the Bureau of Labor Statistics. The latter is based on a quite expensive sample of 60,000 households and comes with a time-lag of one month, while Google trends data is available for free and in real-time (see also [60]).

3. Agent-based simulations for development

Back in 1948, Warren Weaver [63] argued that “science before 1900 was largely concerned with two-variable problems… [while] subsequent to 1900… scientists… developed powerful techniques of probability theory and of statistical mechanics” to analyze a large number of variables. Weaver concluded that there is a third group of problems that deals “simultaneously with a sizable number of factors which are interrelated into an organic whole… [and] cannot be handled with the statistical techniques so effective in describing average behavior… These new problems, and the future of the world depends on many of them, requires science to make a third great advance”. Weaver already predicted that this new way of doing science will be “effectively instrumented by huge computers” [63]. The final breakthrough only came with the digital revolution that enabled computational simulations of these kinds of complex systems some five to six decades later.

3.1. Characteristics of agent-based models

While the first computer simulations of social systems in the 1960s were inspired by existing differential equations and simulated macro-level factors [64], the next decades saw an increasing transition of simulations “from factors to actors” [65]. Computer facilitated methods allow for studying the emergence of non-linear macro patterns that arise out of a multiplicity of dynamical micro interactions [11-14]. The very concept of development, which as a total is more than the sum of its parts, is an example of such social emergence. So-called agent-based models, ABM (also multi-agent models) consist of computer simulations of social systems under constraints. The basic set up is reminiscent of popular videogames, such as SimCity, but is executed with the same scientific rigor an engineer employs to computer-simulate the stability of bridges or earthquake-prone skyscrapers. This involves the specification of a specific number of parameters and variables (time, the number of agents, agents’ attributes and behavior, technological and institutional environment, communication and cooperation among agents, etc.), and the subsequent exploration of arising interaction effects among them [64].

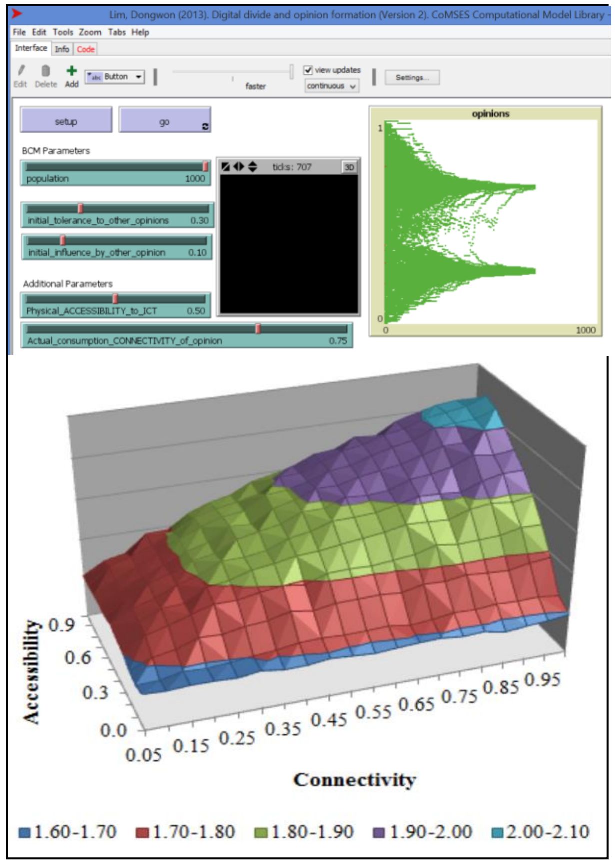


Figure 3: (a) ABM NetLogo implementation of [68]; **(b)** Number of major opinion clusters by accessibility and connectivity [66].

For example, Lim et al. [66] followed a typical ABM setup to simulate the dynamics of opinion formation and fragmentation in a setting of the digital divide. They define the digital divide in terms of “accessibility” (physical access to ICT) and what they call “connectivity” (actual consumption of an opinion). The authors include two additional variables that measure the extent to which individuals tolerate and are influenced by other opinions. Figure 3a shows the implementation of this simple model and its four adjustable variables (plus population size) in Netlogo, a common ABM modelling software [67]. Figure 3b shows that resulting opinion clusters are more fragmented as both accessibility and connectivity increase. This shows that digital connectedness fosters opinion plurality. It also shows that accessibility contributes differentially more to opinion fragmentation as connectivity increases. The authors underline that this can "can keep a society from reaching a consensus" and also show that "extreme opinions better survive in an online environment" [66]. The model additionally allows to quantify how easily a disconnected agent with a novel idea might be isolated, and how quick connected individuals can find other like-minded people to rally support for the idea on the internet.

While this example works with maximal 1000 agents and four adjustable variables (see Figure 3a), there is no conceptual constraint to the number of variables included in the model. Each of millions of individuals can be modelled as a unique case. For example, the spread of infectious disease has been simulated over realistic social networks of California's 39 million inhabitants (including 182 million social network ties) [69]. Each individual was characterized with up to 163 demographic variables from census data. 163 variables give a minimal combinatorial space of 10^49 choices, which is of course more than enough to characterize each individual with a unique profile. The computational power and respective algorithms are already developed to scale this up to the global dimensions. Back in 2011, simulating a global population of 6.57 billion agents with 2.40 billion infections took less than 8 hours of computation time [70].

Formally, agents can be defined as computational entities, usually showing some form of bounded rationality (memory loss, nearsightedness, local search), situated in some environment, capable of undertaking flexible autonomous actions with the objective of satisfying their individual or collective need. The involved variables of the agent and the environment can be fixed or variable (i.e. changing under changing circumstances), which results in a myriad of differential patterns. The result of computer simulations are numerical, which means that they are no deterministic predictions, but a variety of outcomes with different probabilities. This is the result of path-dependency on initial conditions and the incorporation of 'luck' and 'random choice' in the unfolding of a social dynamic, both realistic characteristics of social systems.

3.2. ABMs and development

Agent-based models have several advantages that are useful when applied development work.

3.2.1. 'What-if'? One important aspect is the flexible testing of policy scenarios. In contrary to statistical (big data) analysis that is fixated on a specific past, ABMs are an excellent tool to answer 'what-if' questions. Simulations can test realities that never existed and policies that were never implemented. This is especially important in the social sciences, since (in contrary to the natural sciences, engineering or psychology), it is not an option to sacrifice or manipulate societies in labs. Artificial societies, however, can be manipulated at will without the approval of ethical review boards.

For example, an ABM of 2009 H1N1 outbreak in Mexico showed that government mobility restrictions reduced the spread of the virus by about 10% and postponed it by about 2 days [71]. This policy is in agreement with the recommendation of the World Health Organization, which calls for the suspension of activities in educational, government and business units in case of a pandemic. However, such curfews could cost billions. Agent-based models not can only help to quantify the effect of these kinds of policies, but also simulate scenarios with changed variables, such as (even non-linear) effects of the intensity of restrictions (e.g. closing only airports and not schools), or the use of alternative policies. For example, a simulation of the city Portland has shown that in case of inhalable plague, voluntary mass use of rapidly available antibiotics is as effective as contact tracing, school and city closures [72,73].

Another example from the field of ICTD is the study of the effect of ICT connectivity on collective action (such as in social protests). ABM can explore hypothetical effects of variables that differ among societies, such as online communication patterns and the distribution of political preferences. An ABM demonstrated that the positive role of ICT in both the level and speed of collective actions is not automatic (as often assumed in so-called "Twitter revolutions" [74]), but that it greatly depends on and is highly sensitive to the dispersion of participation preference [75]. This suggests that the effect of ICT in collective action is quite different in contexts with dissimilar preference structures among the involved parties.

3.2.2. Scalable context dependency: Another advantage of ABMs is their modular flexibility, which provides scalable solutions to focus on concrete problems in specific settings, instead of trying to understand general theoretic tendencies [76]. "This is moving from a general theory which is supposed to be applicable everywhere to very context specific models. Such models can for sure share some common bases but they should also be adapted to the specific context." [77]. Reusing the code allows to create tailor-made models for concrete problems in specific, local- and context-dependent settings. Development landscapes are notoriously heterogeneous, and embracing this diversity can be key to understand subtleties of any intervention.

For example, several ABMs exist to investigate civil violence and riots by simulating the contagious nature of spreading participation and the differential consequences of varying intensities of police presence and reaction time [78-80]. Fine-tuned and extended by empirical data, the basic idea behind such models has then been applied to the specific case of the 2011 London riots, which resulted in several deaths and USD 400 million in damages [81]. Similar to the creation of different versions of SimCity [82], the computer simulation of a unique local community can make use of existing software modules, while evaluating a context-dependent future that is different from the past. This provides a cost-effective solution to eventually replace research on 'the representative village in Africa' with 'this specific village in Africa'.

3.2.3. Intuitive science communication: Sticking to the image of SimCity reveals an additional benefit. The multimedia visualization can be used to communicate with, engage, and convince policy makers and stakeholders who lack sophisticated statistical or scientific training. While the use of simulation software programs like TRANSIMS in the late 1990s were quite sterile [73], the application of modern simulations are much more visually rich (see Figure 4) and can be run on an affordable laptop. In contrary to the intimidating equations and static graphs of traditional analysis, the presentation of dynamic computer simulations is as intuitive as watching and playing a videogame and allows for a rather playful approach to development analytics. Stakeholders can see the social dynamic unfolding and even take ownership of the model by asking for real-time adjustment of parameters in order to test for specific scenarios. Policy-makers can test countless 'what if' scenarios on a concrete setting before taking the plunge for one or the other option.

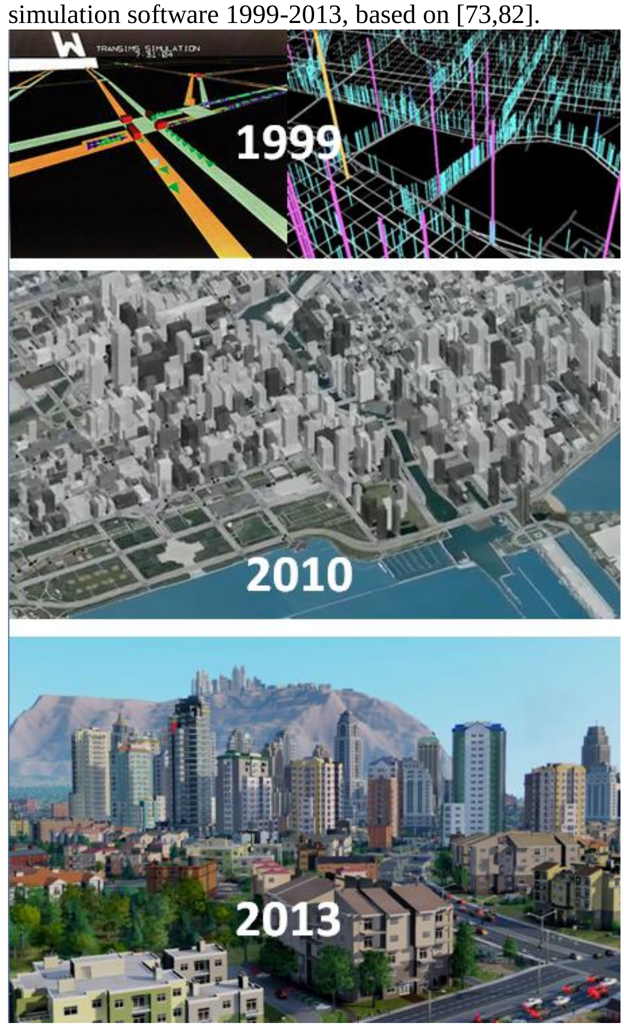
4. Opportunities and Challenges for development

The full potential of computational social science becomes clear when combining both the empirical and the modelling approach. The challenges become clear when remembering that these innovations are subject to the well-known innovation processes of Schumpeterian creative destruction, including its diffusion and learning curves.

4.1. Opportunities

The importance of combining statistical analytics and theory-driven models in the field of development arises from Lucas critique [8]. Data from the past have a limited value after an intervention that is purposefully designed to systematically alter the *modus operandi* of the targeted system. It is important to qualify this statement by pointing out that the validity of data insights depends on the kind of statistical analysis. Some tests are designed to be sufficiently broad to predict a large and general group of cases (i.e. testing for 'out-of-of-group' samples from a more general groups of situations), while others are fitted to explain particular cases [83]. The difference consists in identifying patterns contained in a larger, more general class of cases, and in explaining the particular circumstances of a specific (class of) case. The vast majority of current social science research focuses on the latter kind of explanatory analysis (mostly executed through significance and R^2 tests) [83,84].

Figure 4: Evolution of socio-economic simulation software 1999-2013, based on [73,82].



However, even statistical test that test for a very broad family of cases reach their limit when the structure of the system is altered and no data exists for such changed system. The fine print of social science generally recognizes this fact with the all-pervasive *ceteris paribus* qualification ('all other things being equal'), which seeks to safeguard against the application of the obtained results to cases which are somewhat different from the analyzed case. Lucas critique says that this is contradictory to the goal of any policy, which is aimed at changing the conditions of the analyzed case. The heterogeneity of development contexts provides for the fact that most of the time no valid data exists from such changed system. This limits extrapolation. A developing Africa is not simply an extrapolated version of Europe's past development trajectory, and a connected favela in eastern Brazil is not equal to one in the west.

One alternative offered by the real-time nature of many big data sources is so-called 'nowcasting' [85], the possibility (and actual need [18]) for continuous adjustment of the model with real-time data releases. But this only allows to monitor a policy, not to evaluate policy alternatives beforehand. Adjustable simulation models allow for the exploration of futures that have never been, and data of the specific case allows to adjust the model to the specific context.

In this sense the final goal of computational social science is to combine big data approaches with computer-facilitated modelling techniques. "A good complex systems model both begins and ends with data: Low level data is used to formulate the assumptions about the building blocks of the model, and both and high and low level data is also used to test whether the resulting emergent phenomena properly correspond to those observed in the real world" [86].

Often ABMs are calibrated with coarse-grained records (such as done when modeling the process of knowledge diffusion in Santiago de Chile [87]), or with small-scale survey data, such as done by Wei et al. [88], who collected some 225 questionnaires to calibrate an ABM that simulates the optimization of m-banking adoption. Simulation models from the natural sciences painstakingly collect data input with sensors and cameras, such as the ambitious Madingley model of the world's ecosystem that aims at simulating "all life on earth" [102]. Ecologists report that "the biggest stumbling block… is obtaining the data to parameterize and validate" the model. On the contrary, social scientists do not require "motion-activated cameras… [or] continuous plankton recorders towed beneath ships" to obtain their big data [102]. Humans conveniently produce relevant data as a byproduct of their digital life. Examples include the above-mentioned use of mobility patterns extracted from mobile phone call records to simulate the 2009 H1N1 outbreak in Mexico and to evaluate the impact that government policies had on the spreading of the virus [71], and the use of policy records to simulate the 2011 civil unrest in London [81]. This allows for a "more realistic representation of human behavior which includes the behavioral changes that might take place" during the dynamic under study [71;p.2].

A quite advanced example is the virtual simulation of the city of the U.S. city of Portland [72,73]. This simulation included the modeling of 1.6 million residents with real socio-demographic profiles following identified daily activities in 180,000 specific locations. The data was obtained through a combination of traditional census data, digitally recorded (big data) records and personal activity logs. The results led to surprising insights into complex social dynamics. For example, the city tested for the installation of a new light-rail system (how would traffic patterns and individual behavior change for different rail routes?), and saved millions of US$ as a result. Extensive infrastructure projects like these are also often common in ICTD.

Summing up in the language of economists, one can say that the approach of computational social science "is intermediate between traditional economic theory and econometrics. Traditional economic theory is top-down, modeling decision making from first principles, and then testing against data later… Econometrics, in contrast, takes a bottom up, data-driven, but fundamentally ad hoc approach. The complex systems approach sits in the middle, taking a bottom up data-driven approach that differs from traditional econometrics by explicitly representing agents and institutions and modeling their interactions, without the requirement that everything be derived from fundamental principles" [86].

4.2. Challenges

The transition to computational social science does not happen in a vacuum, but within the structural, institutional, economic, and social reality of existing societies. This affects the potential benefits stemming from it. The rejection of technological determinism dictates that the application of a new technology will never be inherently and automatically good [89]. Emblematic is the case of the digitization of twenty million land records in Bangalore, which created a big data source aimed at benefiting 7 million small farmers in over

27,000 villages [90]. Contrary to expectations, the usual large players were in a much better position to exploit the provided data, resulting in a perpetuation of existing inequalities [91]. Even well-intended practices can turn out the wrong way. Data is power, and power benefits from data.

Most of the limitations of the use of computational social science are very similar to the challenges tackled by traditional ICTD projects, including challenges in the areas of infrastructure, human resources, and institutional frameworks [19].

4.2.1. Access challenges: First and foremost, the lack of infrastructure access and usage limits the availability of any digital footprint. “Twitter does not represent ‘all people’, and it is an error to assume ‘people’ and ‘Twitter users’ are synonymous [92, p. 669]. Not surprisingly, it turns out that the question of sample representativeness is closely linked to the degree of digital inequality. Using our previous terms, the question is how how close the big data *sampling n* gets to the *universe N*. Blumenstock et al. [23,26,35] worked with mobile phone data from Rwanda from 2005-2009, during which the mobile phone penetration was between 2 % and 20 %. They found that “phones are disproportionately owned and used by the privileged strata of Rwandan society” [23]. Frias-Martinez et al. worked with mobile phone big data from a more advanced “emerging economy in from Latin America” [24], with a mobile phone penetration of around 60-80%. The big data sample matched the social stratification of the available census data impressively well.

However, even once everybody is connected to a mobile phone, the continuous bandwidth divide [93] leads to the fact that we will always have better big data sources from some parts of society, but not from others. Over the past two decades, telecom access has ever become more diversified. In the analog age of the late 1980s, the vast majority of telecom subscriptions were fixed-line phones, and all of them had the same performance (see Figure 5). Twenty years later, there’s a myriad of different telecom subscriptions with the most diverse range of performances. Far from being closed, the digital divide incessantly evolves through an ever changing heterogeneous collection of telecom bandwidth capacities [93].

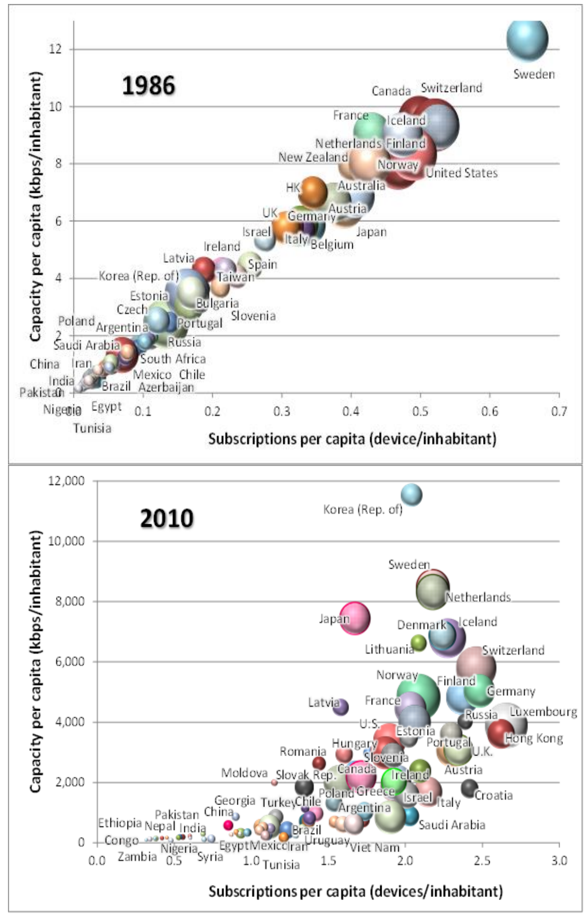
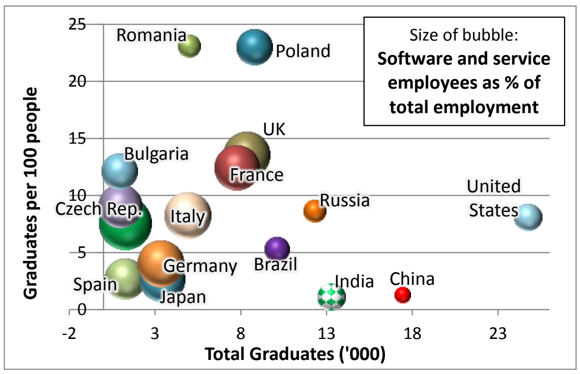


Figure 5: Subscriptions per capita vs. installed bandwidth per capita (in optimally compressed kbps) for 1986 and 2010. Size of the bubbles represents Gross National Income (GNI) per capita (N = 100); based on [93].

4.2.2. Human Resource and skills challenges: Hal Varian, chief economist at Google, emblematically stated: “the sexy job in the next 10 years will be statisticians… And I’m not kidding” [94]. The same counts from programmers and computer scientists. Case studies on the use of big data applications in development projects show that adequate training for data specialists and managers is one of the main reasons for failure [95]. It is predicted that in the near future even the job magnet United States will face a shortage of some 160,000 professionals with deep analytical skills (of a total of 450,000 in demand), as well as a shortage of 1.5 million data managers that are able to make informed decisions based on analytic findings (of a total of 4 million in demand) [3].

Within this context of global shortage, from a relative standpoint of international comparison, Figure 6 shows that some developing countries achieve relatively high graduation rates for professionals with deep analytical skills (high up on the vertical y-axis in Figure 8). In general, countries from the former Soviet bloc (e.g. Romania, Poland, and Bulgaria) produce a high number of analysts. The world's large developing BRIC countries (Brazil, Russia, India and China) produce 40 % of the global professionals with deep analytical skills, twice and many as the university power-hose of the United States (x-axis in Figure 6). They also are expected to have large future demand, as their current share of software and service employees as percentage of total employment is still relatively low (size of bubbles in Figure 6). Traditional leaders of the global economy, such as Germany and Japan, are comparatively ill-prepared to sustain their demand from national training sources.

Figure 6: 2011 graduates with deep analytical training: total (horizontal x-axis), per 100 people (vertical y-axis); software and service employees as % of total employment (size of bubbles); based on [3,20,96]



Not only the quantity, but also the quality of analytical training matters. The inventory of big data social media studies by Kalampokis et al. [41] revealed that more than one third of the exercises that claimed to demonstrate the predictive power of social media did not even run any explicit predictive analytics (but mere explanatory statistics, such as R^2 analysis). This a better ratio than in traditional (non big data) studies (in traditional information systems research it has been shown that merely 13 % of studies that proclaim predictive power use actual predictive analytics [84]), but still shows systematic misuse of statistical techniques in the social sciences.

4.2.3. Institution-building challenges: Last but not least, the ICTD community is well aware that any general-purpose technology revolution also requires adjustments in the corresponding institutions. Schumpeter's creative destruction makes important aspects of previous institutional settings obsolete, while it requires the social construction of new institutions that are up for the digital challenges [9,97]. Privacy concerns, social discrimination and related abuses are among the biggest threats to the application of techniques like those promoted by big data. In a 2014 White House report on big data of office of President Obama underlined that big data leads to "vexing issues (big data technologies can cause societal harms beyond damages to privacy, such as discrimination against individuals and groups)", while at the same time emphasizing the "tremendous opportunities these technologies offer to improve public services, grow the economy, and improve the health and safety of our communities" [99]. The challenge to build institutions that minimize the risks and maximize the benefits is especially delicate in developing countries, in which institutional frameworks are notoriously weak, but the catch-up potential is extraordinarily high.

5. Fostering the use of ICT in ICTD

In the words of fifteen leading scholars in the field: "computational social science is occurring—in Internet companies such as Google and Yahoo, and in government agencies such as the U.S. National Security Agency. Computational social science could become the exclusive domain of private companies and government agencies. Alternatively, there might emerge a privileged set of academic researchers presiding over private data… Neither scenario will serve the long-term public interest" [1]. The goal has to be to take advantage of the ensuing possibilities for very concrete development projects and policy interventions that benefit the general public. This will not only require the contributions of social scientists, but also the adequate design of information systems that consider and integrate such approaches as an inherent part of their design. For example, information systems could be designed in a way that automatically produce privacy respecting big data sources that feed computer simulations which enable to adjust the policy intervention on the go.

The field of ICTD seems very appropriate for such work. On the one hand it counts with a community that is not only convinced of the power of ICT, but also counts with a critical mass of programming and information systems design skills. One in three papers in the field of ICTD come from computer science and engineering [100]. On the other hand, ICTD projects inevitably involve ICT, which will produce big data footprints.

One ongoing challenge consists in significantly increasing the effort to create theory-driven models that can be used in an ever-changing reality. The review in this article has shown that man more examples and applications are available in the field of big data, as compared to the field of computer simulations. Much more effort has to be put into such theory-driven models. If not, we run the risk of falling into the same traps that Lucas' colleagues did in the 1970s, some four decades before the big data revolution.

6. References

[1] Lazer, D. et al. Computational Social. Science 323, 721–723 (2009).

[2] Mayer-Schönberger, V. & Cukier, K. Big data: a revolution that will transform how we live, work and think. (John Murray, 2013).

[3] Manyika, J. et al. Big data: The next frontier for innovation, competition, and productivity. (McKinsey & Company, 2011).

[4] Kolb, J. & Kolb, J. The Big data Revolution. (CreateSpace Independent Publishing Platform, 2013).

[5] Letouzé, E. Big data for Development: Opportunities and Challenges. (United Nations Global Pulse, 2012). at <http://www.unglobalpulse.org/projects/BigDataforDevelopment>

[6] WEF (World Economic Forum) & Vital Wave Consulting. Big data, Big Impact: New Possibilities for International Development. WEF (2012).

[7] Hilbert, M. Big data for Development: From Information - to Knowledge Societies. (Social Science Research Network, 2013). at <http://papers.ssrn.com/abstract=2205145>

[8] Lucas Jr, R. E. Econometric policy evaluation: A critique. Carnegie-Rochester Conference Series on Public Policy 1, 19–46 (1976).

[9] Goodhart, C. Problems of Monetary Management: The UK Experience. (The Public Affairs Center, Dartmouth College, 1976).

[10] Campbell, D. T. Assessing the Impact of Planned Social Change. Occasional Paper Series, #8. (1976). at <http://eric.ed.gov/?id=ED303512>

[11] Schelling, T. C. Micromotives and Macrobehavior. (W. W. Norton & Company, 2006).

[12] Epstein, J. M. & Axtell, R. L. Growing Artificial Societies: Social Science from the Bottom Up. (A Bradford Book, 1996).

[13] Miller, J. H. & Page, S. E. Complex adaptive systems. (Princeton University Press, 2007).

[14] Hummon, N. P. & Fararo, T. J. The emergence of computational sociology. The Journal of Mathematical Sociology 20, 79–87 (1995).

[15] OECD (Organisation for Economic Co-operation and Development). in Supporting Investment in Knowledge Capital, Growth and Innovation 319–356 (Organisation for Economic Co-operation and Development, 2013).

[16] UN Statistical Commission. Big data and modernization of statistical systems: Report of the Secretary-General. (United Nations Economic and Social Council, 2014). at <http://unstats.un.org/unsd/statcom/doc14/2014-11-BigData-E.pdf>

[17] Ginsberg, J. et al. Detecting influenza epidemics using search engine query data. Nature 457, 1012–1014 (2009).

[18] Lazer, D., Kennedy, R., King, G. & Vespignani, A. The Parable of Google Flu: Traps in Big data Analysis. Science 343, 1203–1205 (2014).

[19] Hilbert, M. Big Data for Development: A Systematic Review of Promises and Challenges. (forthcoming).

[20] ITU (International Telecommunication Union). World Telecommunication/ICT Indicators Database. (International Telecommunication Union, 2014).

[21] Naef, E. et al. Using Mobile Data for Development. (Cartesian and Bill & Melinda Gates Foundation, 2014). at <https://docs.gatesfoundation.org/Documents/Using%20Mobile%20Data%20for%20Development.pdf>

[22] Raento, M., Oulasvirta, A. & Eagle, N. Smartphones: An Emerging Tool for Social Scientists. Sociological Methods & Research 37, 426–454 (2009).

[23] Blumenstock, J. E. & Eagle, N. Divided We Call: Disparities in Access and Use of Mobile Phones in Rwanda. Information Technologies & International Development 8, pp. 1–16 (2012).

[24] Frias-Martinez, V. & Virseda, J. Cell Phone Analytics: Scaling Human Behavior Studies into the Millions. Information Technologies & International Development 9, pp. 35–50 (2013).

[25] Martínez, E. F. & Martínez, V. F. Method, computer programs and a use for the prediction of the socioeconomic level of a region. (2014). at <http://www.google.com/patents/US20140032448>

[26] Blumenstock, J. E., Gillick, D. & Eagle, N. Who's Calling? Demographics of Mobile Phone Use in Rwanda. in AAAI Spring Symposium: Artificial Intelligence for Development (2010).

[27] Frias-Martinez, V., Frias-Martinez, E. & Oliver, N. A Gender-centric Analysis of Calling Behavior in a Developing Economy Using Call Detail Records. (2010).

[28] Soto, V., Frias-Martinez, V., Virseda, J. & Frias-Martinez, E. in User Modeling, Adaption and Personalization (eds. Konstan, J. A., Conejo, R., Marzo, J. L. & Oliver, N.) 377–388 (Springer Berlin Heidelberg, 2011). at <http://link.springer.com/chapter/10.1007/978-3-642-22362-4_35>

[29] Hilbert, M. When is Cheap, Cheap Enough to Bridge the Digital Divide? Modeling Income Related Structural Challenges of Technology Diffusion in Latin America. World Development 38, 756–770 (2010).

[30] Buckee, C. O., Wesolowski, A., Eagle, N., Hansen, E. & Snow, R. W. Mobile phones and malaria: modeling human and parasite travel. Travel Med Infect Dis 11, 15–22 (2013).

[31] Wesolowski, A. et al. Quantifying travel behavior for infectious disease research: a comparison of data from surveys and mobile phones. Sci. Rep. 4, (2014).

[32] Bengtsson, L., Lu, X., Thorson, A., Garfield, R. & von Schreeb, J. Improved Response to Disasters and Outbreaks by Tracking Population Movements with Mobile Phone Network Data: A Post-Earthquake Geospatial Study in Haiti. PLoS Med 8, e1001083 (2011).

[33] Lu, X., Bengtsson, L. & Holme, P. Predictability of population displacement after the 2010 Haiti earthquake. PNAS 109, 11576–11581 (2012).

[34] Moumni, B., Frias-Martinez, V. & Frias-Martinez, E. Characterizing Social Response to Urban Earthquakes

Using Cell-phone Network Data: The 2012 Oaxaca Earthquake. in Proceedings of the 2013 ACM Conference on Pervasive and Ubiquitous Computing Adjunct Publication 1199–1208 (2013). doi:10.1145/2494091.2497350

[35] Blumenstock, J., Eagle, N. & Fafchamps, M. Risk and Reciprocity Over the Mobile Phone Network: Evidence from Rwanda. (2012). at <http://archive.nyu.edu/handle/2451/31441>

[36] Telefonica. (2012). Smart Steps. http://dynamicinsights.telefonica.com/488/smart-steps

[37] Statista. Statistics and Market Data on Mobile Internet & Apps. (2014). at <www.statista.com/markets/424/topic/538/mobile-internet-apps/>

[38] Mocanu, D. et al. The Twitter of Babel: Mapping World Languages through Microblogging Platforms. PLoS ONE 8, e61981 (2013).

[39] Frias-Martinez, V. & Frias-Martinez, E. Spectral clustering for sensing urban land use using Twitter activity. Engineering Applications of Artificial Intelligence 35, 237–245 (2014).

[40] Chunara, R., Andrews, J. R. & Brownstein, J. S. Social and news media enable estimation of epidemiological patterns early in the 2010 Haitian cholera outbreak. Am. J. Trop. Med. Hyg. 86, 39–45 (2012).

[41] Kalampokis, E., Tambouris, E. & Tarabanis, K. Understanding the predictive power of social media. Internet Research 23, 544–559 (2013).

[42] Hanneman, R. & Riddle, M. Introduction to social network methods. (University of California, Riverside). at <http://www.faculty.ucr.edu/~hanneman/nettext/>

[43] Newman, M. Networks: An Introduction. (Oxford University Press, USA, 2010).

[44] Eagle, N., Macy, M. & Claxton, R. Network Diversity and Economic Development. Science 328, 1029–1031 (2010).

[45] Hidalgo, C. A., Klinger, B., Barabási, A.-L. & Hausmann, R. The Product Space Conditions the Development of Nations. Science 317, 482–487 (2007).

[46] Valente, T. W. & Saba, W. P. Mass media and interpersonal influence in a reproductive health communication campaign in Bolivia. Communication Research 25, 96 (1998).

[47] Valente, T. W. Network Models of the Diffusion of Innovations. (Hampton Press (NJ), 1995).

[48] Vishwanath, A. & Barnett, G. A. The Diffusion of Innovations: A Communication Science Perspective. (Peter Lang International Academic Publishers, 2011).

[49] Valente, T. W. & Davis, R. L. Accelerating the Diffusion of Innovations Using Opinion Leaders. The Annals of the American Academy of Political and Social Science 566, 55–67 (1999).

[50] Banerjee, A., Chandrasekhar, A. G., Duflo, E. & Jackson, M. O. The Diffusion of Microfinance. (National Bureau of Economic Research, 2012).

[51] Lee, S. & Monge, P. The Coevolution of Multiplex Communication Networks in Organizational Communities. Journal of Communication 61, 758–779 (2011).

[52] Baron, L. F. & Gomez, R. Social Network Analysis of Public Access Computing: Relationships As a Critical Benefit of Libraries, Telecenters and Cybercafes in Developing Countries. in Proceedings of the 2012 iConference 377–383 (ACM, 2012).

[53] UNDP (United Nations Development Programme). Human Development Index; http://hdr.undp.org

[54] Nobel Prize. The Sveriges Riksbank Prize in Economic Sciences in Memory of Alfred Nobel 1998: Amartya Sen (1998)

[55] Stiglitz, J., Sen, A. & Fitoussi, J.-P. The Measurement of Economic Performance and Social Progress Revisited. (Commission on the Measurement of Economic Performance and Social Progress, 2009).

[56] Frey, B. S. & Stutzer, A. Happiness and Economics: How the Economy and Institutions Affect Human Well-Being. (Princeton University Press, 2010).

[57] MarketPsych. Thomson Reuters MarketPsych Indices (TRMI); https://www.marketpsych.com/data/

[58] Anderson, C. The End of Theory: The Data Deluge Makes the Scientific Method Obsolete. Wired Magazine (2008).

[59] Althouse, B. M., Ng, Y. Y. & Cummings, D. A. T. Prediction of Dengue Incidence Using Search Query Surveillance. PLoS Negl Trop Dis 5, e1258 (2011).

[60] Ettredge, M., Gerdes, J. & Karuga, G. Using Web-based Search Data to Predict Macroeconomic Statistics. Commun. ACM 48, 87–92 (2005).

[61] Google correlate dengue trends, http://www.google.org/denguetrends/about/how.html

[62] Hubbard, D. W. Pulse: The New Science of Harnessing Internet Buzz to Track Threats and Opportunities. (Wiley, 2011).

[63] Weaver, W. Science and Complexity. American Scientist 36, 536–544 (1948).

[64] Gilbert, N. & Troitzsch, K. Simulation for the Social Scientist. (Open University Press, 2005).

[65] Macy, M. W. & Willer, R. From Factors to Actors: Computational Sociology and Agent-Based Modeling. Annual Review of Sociology 28, 143–166 (2002).

[66] Lim, D., Lee, H., Zo, H. & Ciganek, A. Opinion Formation in the Digital Divide. JASSS 17, 13 (2014).

[67] Wilensky, U. NetLogo. (The Center for Connected Learning (CCL) and Computer-Based Modeling, 1999). at <https://ccl.northwestern.edu/netlogo/>

[68] Lim, D. Digital divide and opinion formation (Version 2). CoMSES Computational Model Library (2012). at <http://www.openabm.org/model/3361/version/2/view>

[69] Barrett, C. L., Bisset, K. R., Eubank, S. G., Feng, X. & Marathe, M. V. EpiSimdemics: An efficient algorithm for simulating the spread of infectious disease over large realistic social networks. in High Perf.Comp. (2008).

[70] Parker, J. & Epstein, J. M. A Distributed Platform for Global-Scale Agent-Based Models of Disease

Transmission. ACM Trans Model Comput Simul 22, 2 (2011).

[71] Frias-Martinez, E., Williamson, G. & Frias-Martinez, V. An Agent-Based Model of Epidemic Spread Using Human Mobility and Social Network Information. in 2011 IEEE Third Intern. Conf. on Social Computing 57–64 (2011). doi:10.1109/PASSAT/SocialCom.2011.142

[72] Barrett, C., Eubank, S., & Smith, J. (2005). If Smallpox Strikes Portland... Scientific American, 292(3), 54–61.

[73] Mac Hyman (2012). Part 6 Simulation Models: TRANSISMS and EpiSims. Disaster Resillence leadership Academy“; http://www.youtube.com/watch?v=pGftX_56X8g

[74] Morozov, E. Moldova’s Twitter Revolution. Foreign Policy Blogs (2009).

[75] Hu, H., Cui, W., Lin, J. & Qian, Y. ICTs, Social Connectivity, and Collective Action: A Cultural-Political Perspective. JASSS 17, 7 (2014).

[76] Banerjee, A., Banerjee, A. V. & Duflo, E. Poor Economics: A Radical Rethinking of the Way to Fight Global Poverty. (PublicAffairs, 2011).

[77] Wendelspiess, F. Agent-Based Models in Development Economics. (SSRN, 2014). at <http://papers.ssrn.com/abstract=2397252>

[78] Epstein, J. M. Modeling civil violence: An agent-based computational approach. PNAS 99, 7243–7250 (2002).

[79] Yiu, Sau Yee, Andrew Gill, and Peng Shi. Investigating strategies for managing civil violence using the MANA agent based distillation. Land Warfare Conference, Brisbane. 2002.

[80] Goh, C. K., Quek, H. Y., Tan, K. C. & Abbass, H. . Modeling Civil Violence: An Evolutionary Multi-Agent, Game Theoretic Approach. in IEEE Congress on Evolutionary Computation, 2006. CEC 2006 1624–1631 (2006). doi:10.1109/CEC.2006.1688503

[81] Davies, T. P., Fry, H. M., Wilson, A. G. & Bishop, S. R. A mathematical model of the London riots and their policing. *Sci. Rep*. 3, (2013).

[82] Electronic Arts. SimCity Official Website. (2014). at <http://www.simcity.com/>

[83] Shmueli, G. To Explain or to Predict? Statistical Science 25, 289–310 (2010).

[84] Shmueli, G. & Koppius, O. R. Predictive Analytics in Information Systems Research. MIS Q. 35, 553–572 (2011).

[85] Giannone, D., Reichlin, L. & Small, D. Nowcasting: The real-time informational content of macroeconomic data. Journal of Monetary Economics 55, 665–676 (2008).

[86] Farmer, J. D. Economics needs to treat the economy as a complex system. (Complexity Research Initiative for Systemic instabilities (CRISIS), 2012).

[87] Piergiuseppe, M. & Taylor, R. Small World Dynamics and The Process of Knowledge Diffusion: The Case of The Metropolitan Area of Greater Santiago De Chile. JASSS 7, (2004).

[88] Wei, X., Hu, B. & Carley, K. M. Combination of Empirical Study with Qualitative Simulation for Optimization Problem in Mobile Banking Adoption. JASSS 16, 10 (2012).

[89] Kranzberg, M. Technology and History: ‘Kranzberg’s Laws’. Technology and Culture 27, 544 (1986).

[90] Chawla, R. & Bhatnagar, S. Online Delivery of Land Titles to Rural Farmers in Karnataka, India. in (2004). at <http://info.worldbank.org/etools/docs/reducingpoverty/case/96/fullcase/India%20Bhoomi%20Full%20Case.pdf>

[91] Benjamin, S., Bhuvaneswari, R. & Manjunatha, P. R. Bhoomi: ‘E-Governance,’ or, An anti-politics machine necessary to globalize Bangalore. (2007). at <http://casumm.files.wordpress.com/2008/09/bhoomi-e-governance.pdf>

[92] boyd, danah & Crawford, K. Critical Questions for Big Data. Information, Communication & Society 15, 662–679 (2012).

[93] Hilbert, M. Technological information inequality as an incessantly moving target: The redistribution of information and communication capacities between 1986 and 2010. J Assn Inf Sci Tec 65, 821–835 (2014).

[94] Lohr, S. For Today’s Graduate, Just One Word: Statistics. The New York Times (2009). at www.nytimes.com/2009/08/06/technology/06stats.html

[95] Noormohammad, S. F. et al. Changing course to make clinical decision support work in an HIV clinic in Kenya. International J. of Medical Informatics 79, 204–10 (2010).

[96] UNCTAD (United Nations Conference on Trade and Development). Information Economy Report 2012: The Software Industry and Developing Countries. (2012).

[97] Perez, C. in Globalization, Economic Development and Inequality: An alternative Perspective (ed. Reinert, E.) 217–242 (Edward Elgar, 2004). at <http://www.carlotaperez.org/papers/basic-technologicalrevolutionsparadigm.htm>

[98] Hilbert, M. Towards a Conceptual Framework for ICT for Development: Lessons Learned from the Latin American ‘Cube Framework’. Information Technologies & International Development 8, 243–259 (Spanish version: 261–280) (2012).

[99] White House. Big data: Seizing Opportunities, preserving values. (Exceutive Office of the President, 2014). at <http://www.whitehouse.gov/issues/technology/big-data-review>

[100] Hanafizadeh, M. R., Hanafizadeh, P. & Bohlin, E. Digital Divide and e-Readiness: Trends and Gaps. International Journal of E-Adoption 5, 30–75 (2013).

[101] Overeem, A., Leijnse, H. & Uijlenhoet, R. Country-wide rainfall maps from cellular communication networks. PNAS 110, 2741–2745 (2013).

[102] Purves, D. et al. Ecosystems: Time to model all life on Earth. Nature 493, 295–297 (2013).