ICIS 2022 Proceedings

IT Policy and Government

Dec 12th, 12:00 AM

A Social Science Approach Using Big Data for City Planning

Kim Huat Goh
Nanyang Technological University, akhgoh@ntu.edu.sg

Waifong Boh
Nanyang Technological University, awfboh@ntu.edu.sg

Premchand Dommaraju
Nanyang Technological University, premchand@ntu.edu.sg

A Social Science Approach Using Big Data for City Planning

Short Paper

Kim Huat Goh
Nanyang Technological University
Nanyang Business School,
50 Nanyang Avenue, Singapore 39798
akhgoh@ntu.edu.sg

Wai Fong Boh
Nanyang Technological University
Nanyang Business School
50 Nanyang Avenue, Singapore 39798
awfboh@ntu.edu.sg

Premchand Dommaraju
Nanyang Technological University
School of Social Science
48 Nanyang Ave, Singapore 639818
premchand@ntu.edu.sg

Abstract

Cities bring about economic dynamism through positive economic externalities; however, the concentration of people in dense locations has its costs — epidemics, social unrest, pollution, and congestion are some of the ills of cities. As cities evolve, they experience stress, and fault lines appear; the ability to pulse a city and provide early warning of these fault lines can prove advantageous for policymakers in managing and planning for cities. This paper outlines a research program that developed a city scanning tool to measure cities and detect aberrations as they surface. We aggregated data from various industry partners, governmental agencies, and public online sources to develop the measurement metric and applied social science theories to analyze and interpret the results. The results of this study contribute to information system (IS) research by showcasing the role IS research in city planning and for societal good.

Keywords: City planning, big data, social science, outlier detection, city pulse

Introduction

Cities are "man's greatest invention." They make us "richer, smarter, greener, healthier and happier" (Glaeser 2012)

Although cities bring about economic dynamism through positive economic externalities, the concentration of people in dense locations has its costs — epidemics, social unrest, pollution, and congestion are some of the ills associated with cities. As cities evolve, they experience stress, and fault lines appear; the ability to pulse a city and provide early warning of these fault lines can prove advantageous for policymakers in managing and planning for cities.

Cities are complex systems, and in recent years, scholars have argued that cities behave more like living organisms than large mechanistic structures (Batty 2012; Batty 2013). To understand cities for city planning, policymakers need up-to-date information and a well-defined analysis of the city's status. With these data and analyses, policymakers can better understand the current challenges the city faces and possible threats it will encounter in the short, medium, and perhaps long term. The ability to periodically pulse a city is likened to performing regular health checks for a living organism, thereby allowing policymakers to apply timely interventions when managing cities and crafting policy directions for urban design and planning.

In recent years, this informational need for city planning can be partially addressed by the exponential growth of available city-related data. However, this deluge of city data brings new challenges for city planners. A review of analytics-related work in city planning (Giatsoglou et al. 2016; Hasija et al. 2020; Kose et al. 2020; Miranda et al. 2016) revealed three significant challenges in using these data for city planning. First, although plenty of data is available to analyze modern, developed cities, the data sources are often individual silos, making any effort to agglomerate and harmonize the data difficult. The data owners are disparate, ranging from more restrictive government and business sources to more readily accessible user-generated content and publicly available websites; hence, coordinating and assembling the data for city planning is systematically challenging.

Second, despite significant research efforts to develop methodologies to pulse cities (cf. research leading from the CityPulse initiative, funded by the European Union (Barnaghi et al. 2014), many of these studies take an engineering and mechanistic approach towards measuring cities. Recent academic thinking for city planning has advocated a more social and holistic approach (Batty 2012; Batty 2013), calling for a stronger focus on social-technical aspects of cities. Particularly, there is an ongoing argument to transition the thinking of 'cities as machines' to cities as organisms' (Batty 2012) and consider the role of interactions among human agents within the system. Third, when pulsing cities, the issue of residents' privacy is of paramount concern. The ability to measure the condition of a city in real-time and at the same time maintain the anonymity of its residents is critical to ensure acceptance among city's residents.

This research study aims to develop a multi-perspective framework to derive composite measures to detect trends, identify deviations, and predict and validate relationships across different cities. Particularly, the composite measures developed will be used for policymakers to pivot and drill down for greater interpretability and city planning. Additionally, we exhibit the feasibility of this approach by collecting two years of data using this framework and detecting emerging trends in a city. This framework serves as a way for policymakers to pulse a city and track the 'city's health' – analogous to physiological measures such as the heart rate or blood pressure in humans, which can be used to track the health of individuals.

This paper describes a nationally-funded research program across three academic and research institutions and three corporate collaborators. The findings of this research contribute to the literature on urban planning. Particularly, it showcases the role big data can play for city planning by alerting policymakers of potential red flags in a city (e.g. infrastructure congestion or epidemics) while drawing up mid to long-term plans for the city. From a design science perspective, this study highlights a methodology by which one can aggregate and harmonize disparate sources of city data while ensuring the feasibility and practicality of these results for urban planning. Finally, this study shows the significance of social science when interpreting big data for city planning. In particular, we argue in designing a system to measure cities, it is essential to take a socio-technical perspective informed by information systems thinking instead of a completely engineering perspective. The role of information systems research presents useful guiding principles in this development process.

Theoretical Foundations

Various research focusing on measuring cities has been published in the last decade. There are works centered on a whole range of aspects, from providing an understanding of the concept to proposing frameworks and architectures that would support gathering, aggregating, and visualizing various indicators of city pulse (Barnaghi et al. 2014; McKenzie et al. 2015; Miranda et al. 2016; Obaid et al. 2012; Puiu et al. 2016). Out of these, some aim for conceptual clarification (McKenzie et al. 2015), while others focus on semantic models (Bischof et al., 2014) or technical architecture (Puiu et al. 2016). Researchers have also attempted narrowing down specific dimensions of city pulse based on easily trackable factors such as human mobility as well as activities, including attendance of events (Vaccari et al. 2010), biking (Froehlich et al. 2009), spatio-temporal activity (Miranda et al. 2016) and geo-located social activity (Giatsoglou et al. 2016). Much of the research aims to develop various indicators of what could be termed the city pulse through factors such as those that enable monitoring of human mobility or activity, the region's economic health, and social media use. Few have adopted a comprehensive, integrated means of assessing the pulse of a nation aggregating a wide range of factors that may be indicative of the vibrancy of city life.

We performed a comprehensive review of research attempting to measure cities and systematically codify and classify the measures into different themes. 79 articles related to data and city planning were collated; their references are not reported in this paper for brevity. Two researchers read these articles and deemed 45 articles to provide sufficient information about the measures used. Of the measures proposed, the researchers classified them using a grounded approach, and five themes emerged from this process as seen in Table 1 below. This systematic review allows us to ensure a holistic and possibly exhaustive search of the possible measures while presenting an empirically feasible manner of measuring cities.

Dimensions	Societal (S)	Health (H)	Attitudes (A)	Economic (E)	Environmental (E)
Count of articles referencing the dimension	26 out of 45 articles	26 out of 45 articles	5 out of 45 articles	24 out of 45 articles	27 out of 45 articles

Table 1: Dimensions (SHAEE) of Cities as Measured in Prior Literature

While performing the literature review, we have identified that our research extends existing work in three areas. First, extant literature focused only on a few dimensions for city planning and did not have a holistic measurement of different aspects of the city. Second, prior work focuses on relatively macro measurements and/or specific granular data (e.g. traffic flow) as silos. In this work, we seek to harmonize macro indicators and granular data that can or may be used to pulse a city. Third, this paper is among the first to use big data to detect outliers and aberrations within a city.

In essence, the literature review establishes the '*what*' – what city dimensions should we be mindful of, guiding us in sourcing relevant data to measure a city. The next theoretical challenge is the '*how*' – how should one envision the use of this data collected; how will the data collected inform emerging trends in the city; how can policymakers these measures for city planning. To this end, we adopt the thinking of symbolic interactionism as a theoretical base to guide the use of the data collected under this framework.

Symbolic interactionism (SI) is a sociological theory that argues that actions depend on meaning while individuals derive meaning from social interactions, and meaning does change over time through these interactions (Blumer 1969). In the most conventional form, SI is a micro-sociological theory where the unit of analysis is the individual, with the individuals as the agents of interaction. However, more recent work in SI argues that SI can and should be adopted with a more meso and perhaps macro perspective to account for trends observed in communities with a more aggregated unit of analysis (Collins 1988; Turner 2011). In this study, we answer this call and adopt the SI thinking to study communities within cities with a more aggregated perspective. We apply this theory to examine the city's pulse as urban researchers have likened cities to organisms and the interactions within the cities that matter. Additionally, cities are made up of institutions that play a symbolic role in their communities based on the stakeholders they represent (e.g. societal clubs, government agencies, religious venues etc.). The use of SI as a theoretical base provides us with a lens to simulatanously look at how symbols in cities interact amongst one another and how these interactions can be used to pulse a city.

To observe how SI can be used to study communities, we rely on the research work by Snow in social movements and collective behavior (Snow 2001; Snow 2013). Here he argues that to understand societies and collective behavior, there is an imperative need to understand their '[interactions], whether actual, virtual or imagined' (Snow 2001). The emphasis on actual and virtual suggests that the data to be observed should include both data in the physical and digital environments. Snow further argues that SI can be extended by encompassing four principles to guide researchers in using this theory to understand *collective behavior in societies*.

Briefly, the four principles prescribed by Snow are: First, the principle of symbolism – objectives in a community take on a different meaning that evokes feelings and action. Second, the principle of human agency – humans are fettered by societal structures while at the same time they challenge them. Third, the principle of interactive determinism – is the need to study the interaction among social objects to understand them. Finally, the principle of emergence – social changes arising from daily routines' departures help explain emerging trends we see in communities. These significant deviations form the basis of collective behavior. The prescription of these four principles serves as a guiding principle for determining the locus of data collection for pulsing a city, which we will describe in the next section.

Research Framework

To guide our development of the data collection framework, we apply the four principles of SI proposed by Snow in our methodology development. Given that the amount of data that can measure the status of a city is infinitely large, the SI theory is particularly valuable here as it serves to circumscribe the locus of data to be *collected* and how the data should be *analyzed*. We observed that Principles one and two are useful in defining the locus of data to be collected, and Principles three and four provide the theoretical directions on how the data should be analyzed. Table 2 below summarizes how the four Principles of SI guide the collection and analysis of the data in this methodology.

Principle of SI	Operationalization & Application to Data Methodology
Symbolism	Guide data **collection** – guide the location of data to be collected i.e. to focus on data in key points of interest in the city
Human agency	Guide data **collection** – guide the type of data to be collected i.e. focus on activities involving humans (residents)
Interactive determinism	Guide data **analysis** – the need to analyze multiple perspectives for the different dimensions of data collected and the interaction across these dimensions.
Emergence	Guide data **analysis** – what should we be observing i.e. to detect the emergence of trends through the manifestation of outliers in the data (Batty 2008)

Table 2: Principles of Symbolic Interactionism Guiding Methodology Development

Theory guiding the locus for data collection: The first principle of symbolism prescribes that individuals establish meanings towards objects or institutions in their environment. This first principle informs the spatial dimension in our methodology by advocating the focus of data to concentrate on key areas, points of interest, and institutions across the city. Particularly within a city, sociologists have argued that public spaces are critical for people from diverse backgrounds to communicate, interact and derive meaning – a concept coined as 'cosmopolitan canopies' (Anderson 2004). This emphasis on public spaces as symbols suggests that systems designed to measure the city should use these public places as a focal point of data collection. From a practical standpoint, data from public points of interest are also accessible and representative of the resident population where most, if not all, are allowed to visit.

The second principle of human agency further helps to scope the focus of data collection by emphasizing the role of humans' actions and behavior. This informs that all data collected must be related to human actions, thereby excluding a vast set of non-human, action-related data such as weather. These exclusion criteria ensure parsimony in the data collection process and right-size the development of the required data structures.

Theory guiding the approach for data analysis: The third principle of interactive determinism guides the analysis of the data collected by focusing the analysis on the interaction of different dimensions of data collected. The role of interaction across different aspects of the city is essential for one to observe any evolving trends in the city. By focusing on the interactions, emergent collective behavior can be detected and possibly interpreted through the lens of these interactions. Finally, the fourth principle of emergence guides the analysis by highlighting the importance of tracking aberrations and outliers of the data across these interactions. Specifically, Snow argues that emergent behavior in societies can be seen through social interactions and '[novel states], constitute departures from [...] everyday routines, practices, and perspectives (Snow 2001). Hence, the focus of the analysis is the emphasis on departures from the norm, which can be measured as statistical outliers within the data collected – a concept we will discuss next.

In Batty's seminal works uncovering the underlying science of city development (Batty 2008; Batty 2012; Batty 2013), he argues for the idea of scaling in cities. Here, he argues that objects within a city will scale with respect to the physical space e.g., a small town is more likely to have a tram system than a developed subway system. Notably, the concept of object here is nomologically analogous to the symbols prescribed by SI. Additionally, he argues that interactions within the city should also scale to the city's attributes. For example, large cities should have more intensive human interactions than rural towns. The concepts of scale and deviation from scale are also echoed by economists studying city developments in prior research (Gabaix 1999; Rozenfeld et al. 2008). The manifestation of emergence (as argued in SI) can be empirically observed in cities as statistical outliers and aberrations from the scale of the city. This forms the theoretical basis for measuring, isolating, and interpreting the outliers as we observe them in the data collected.

Having established the logic behind our data collection and analysis, we now discuss the types of data to be collected. The comprehensive review of prior literature and our conceptualization of these five dimensions (SHAEE) seen in Table 1 guides us to focus on data along with these perspectives. Given that each type of data is of a different temporal basis, we need to represent the velocity of changes in the pulse of each community. The data collected could range from static data such as geospatial information including maps to semi-dynamic data such as economic growth and healthcare quality indicators which may be subjected to change although at a slower rate. Additionally, we also expect highly dynamic data such as people's movement and behaviors, including driving, shopping, and eating out, to name a few. Finally, the data is captured at high granularity - at the community level, which can be aggregated at different levels of analysis up to the country level. Figure 1 is a diagrammatic depiction of how we organize and integrate the different sources of data to construct the methodology for measuring the city's pulse.

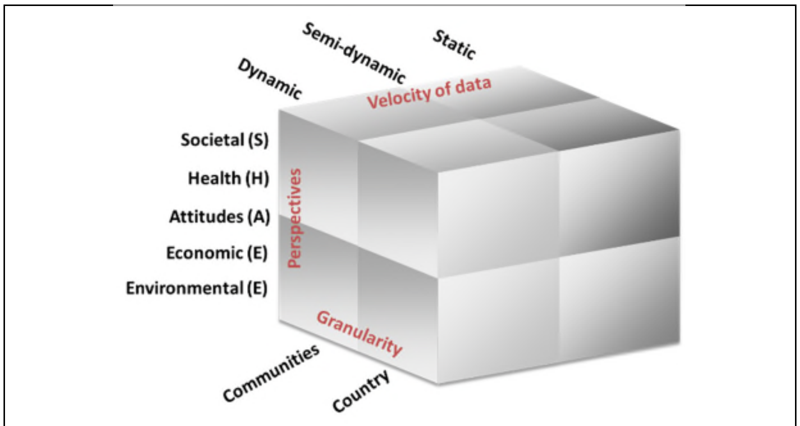

Figure 1. SHAEE – A Multiple Perspective Data Cube

Details of Data Collected

A significant part of this study is to showcase the possibilities and promise in aggregated data from different sources to develop a holistic view of a city. Here, we describe how we aggregate data from *three private organizations* (a healthcare provider, a major credit card network provider, and a mobile telco provider), *various government agencies*, and *social media* / online data from the web. We will also address the issue of privacy concerns at the end of this section.

To collect and aggregate the required data for the five dimensions described earlier, we collaborated with three organizations and multiple government agencies for data access. A summary of the data sources can be seen in Figure 2 below.

To capture economic activity, we obtained credit card payment data from a world-leading credit payment network. The data contains spending patterns (e.g., transaction amount, count, frequency of card use) across 17 different industry segments (e.g., retail, consumer electronics, grocery, etc.) across different communities monthly for a period of 3 years. Here we can observe the credit and debit card payments from both the retailer (revenue) and the consumer (expenditure) perspective, thereby pulsing the 'Economic (E)' aspects of different sub-communities. Additional details of the data collected can also be seen in Table 3.

To capture healthcare activity, we obtained healthcare data from a national electronic health record network that will summarize quarterly the count of residents from different communities hospitalized due to different ailments based on the International Classification of Diseases (ICD-10) classification. In particular, we focus on six major categories of ailments that are more pertinent to city monitoring and planning. They include infectious diseases, chronic illnesses, mental well-being, injuries, substance abuse, and other cause of concern. Each illness category is further broken down into sub-categories to allow the user to drill down into the epidemiology for different communities. For example, infectious diseases are divided into 'infectious and parasitic diseases (ICD-10: A00 to B99)', 'Acute upper respiratory infections (ICD-10: J00 to J06)' etc. Additional details can be seen in Table 3.

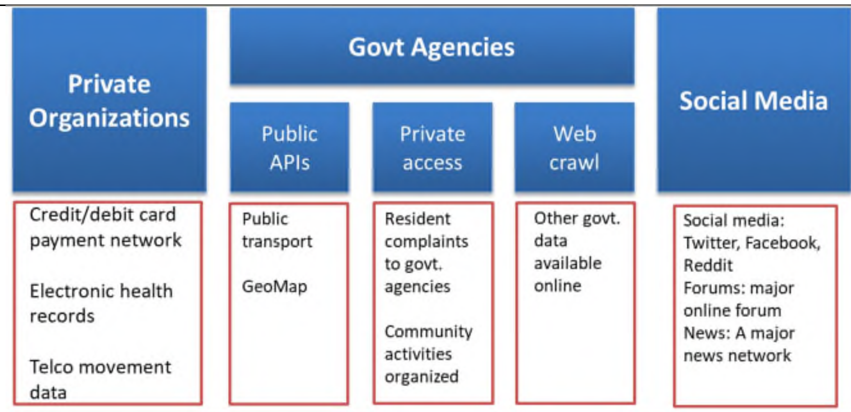

Figure 2: Sources of Data

To capture the environmental movement data, we obtained data from a major telco operator in the country to capture the anonymized movement data of their customer. Here we can measure the stock and flow of individuals in various points of interest, with high-level meta information such as race, gender, and age. We point to major points of interest, including public institutions, major city centers, urban green spaces, etc., to establish human traffic volume at representative times of the month.

In addition to data from organizational sources, we also obtained data from various governmental agencies. We classify these governmental sources into three types based on access to the data. First are public APIs provided by government agencies, including public transportation APIs that provide real-time data on the number of passengers riding on buses, subways, and private taxis. Additionally, we obtain geo-map information from map APIs provided by the state for longitude and latitude coordinates of key places of interest and coordinates of boundaries for different cities. Second, we obtained private access data from government agencies, including formal, written complaints submitted by residents to their city council and participation information in community activities organized by the city council. Third, we also capture all publicly available data provided by the government data portals.

Our third data source comes from social media, which broadly includes data posted by city residents. The sources of social media data collected included Twitter, public pages of Facebook of major public institutions in the city, a major news network website, Reddit, and another popular forum where many residents in the cities post their comments. The text comments and articles of these five major sources are subject to text mining before additional analysis – which we will describe in the next section.

Data (or Variables) Collected	**Perspective (SHAEE)**	**Source**	**Granularity**	**Velocity**
Residential property transacted values (Affluence)	Economic	Government	Community	Daily
Retailer: Txn Amount (by community)	Economic	Payment network provider	Community	Monthly
Retailer: Txn Count (by community)	Economic	Payment network provider	Community	Monthly

Retailer: Avg Ticket Size (by community)	Economic	Payment network provider	Community	Monthly
Number of distinct cards used (by community)	Economic	Payment network provider	Community	Monthly
Avg ticket size per card use (by community)	Economic	Payment network provider	Community	Monthly
Avg frequency of use per card (by community)	Economic	Payment network provider	Community	Monthly
Avg amount spent per card (by community)	Economic	Payment network provider	Community	Monthly
Count of visits (<15 minutes) in key Points of Interest by age, gender, race	Environment	Major Telco	Community	Monthly
Count of stay (>15 minutes) in key Points of Interests by age, gender, race	Environment	Major Telco	Community	Monthly
Traffic speed at major roads	Environment	Government	Community	Real time
Waiting time for public transport	Environment	Government	Community	Real time
Carpark spaces availability	Environment	Government	Community	Real time
Passenger volume by bus stop	Environment	Government	Community	Monthly
Passenger volume by train stations	Environment	Government	Community	Monthly
Taxi availability	Environment	Government	Community	Real time
Public transport ridership	Environment	Government	Community	Monthly
Infectious diseases (further broken down into 5 ICD-10 categories)	Health	Electronic medical records	Community	Quarterly
Chronic diseases (further broken down into 6 ICD-10 categories)	Health	Electronic medical records	Community	Quarterly
Mental illness (further broken down into 5 ICD-10 categories)	Health	Electronic medical records	Community	Quarterly
Injuries (further broken down into 5 ICD-10 categories)	Health	Electronic medical records	Community	Quarterly
substance abuse (further broken down into 2 ICD-10 categories)	Health	Electronic medical records	Community	Quarterly
Other causes of concern: Obesity, malnutrition (further broken down into 2 ICD-10 categories)	Health	Electronic medical records	Community	Quarterly
Lifespan	Health	National statistics	National	Annually
Formal complaints from residents	Societal	Government	Community	Quarterly
Social media comments from Twitter (Geotag to the country)	Societal, Attitudinal	Twitter	Community	Real time
Social media comments from Facebook (Public institution accounts)	Societal, Attitudinal	Facebook	Community	Real time
All news articles from the major news network	Societal, Attitudinal	Major news network	Community	Real time
All news articles from the community news network	Societal, Attitudinal	Community news network	Community	Real time
Relevant threads from Reddit forums	Societal, Attitudinal	Reddit	Community	Real time
All threads from a major online discussion board	Societal, Attitudinal	Major online forum	Community	Real time

Table 3: Data (& Variables) Collected

While collecting the data, maintaining the privacy of the city's residents is of the utmost importance in this study. To ensure this research is translational for practice, the ability to maintain the residents' privacy is critical. The healthcare data collected is fully anonymized according to the Health Insurance Portability and Accountability Act of 1996. Financial payment and mobile users' movement data are geo-aggregated and indexed to ensure no individuals can be identified. Similarly, private data obtained from government agencies were anonymized at the source and aggregated in the reporting process.

Analysis & Early Findings

Some of the data collected are unstructured text for example, social media data and formal complaints to the city council by the residents. All the unstructured data collected are subjected to three forms of text mining. First, we performed Latent Dirichlet Allocation (LDA) to extract the common topics discussed in these textual data. Next, with the LDA-derived topics, we adopted zero-shot classification, specifically BERT, to classify these topics into one of the five dimensions (SHAEE) in our model. Unsupervised text-mining techniques here are required because we are handling a large quantum of unstructured data, and these mining approaches will allow us to summarize these data to collapse into the five dimensions of our measures. Finally, we also performed sentiment analysis to determine the valence of the discussion and social media interactions we observed in these textual comments.

To detect emerging trends within the city and deviations across the cities, we subject all the data collected to outlier analysis. Batty (2008) suggested that empirical phenomena in cities should fall within a particular scale, and deviations from the scale would suggest potential emerging phenomena that may or may not be a cause for concern. Nevertheless, the methodology derived here should raise these aberrations to the attention of policymakers to exercise their city management judgments. To this end, the resultant data cube includes all the natural language processed textual data and all other structured data collected e.g. healthcare, physical movement data, public transport movement data. We performed four forms of outlier analysis, including isolation forests, minimum covariance determinant, local outlier factor, and one-class support vector machine. Observations that are flagged to be outliers in more than two different analyses are singled out for additional review by the research team and possibly to the policymaker for further intervention. To illustrate, below is an example of an aberration detected.

*"We detected a disproportionately high percentage of residents from two communities were hospitalized due to asthma (ICD-10 code: J45) where asthma was the number one reason for admission for these residents. This empirical observation deviates from the general nationwide incidence rate of asthma. (**Principle of Emergence**). Additional analysis of other data collected suggests interactive effects among different data captured **(Principle of Interactive Determinism).** For example, this high asthma incidence correlates with a very high proportion of residents living in rental assistance housing projects where the ventilation of common areas is less than ideal. Additionally, the social-economic status of the communities is significantly below that of the national average, which has lifestyle implications."*

The above example illustrates possible emerging trends within a city, and we caution the reader that we are not arguing any causal relationships leading from this empirical observation. Such a city scanning system aims to *expeditiously* highlight potential aberrations as they surface and not prescribe any causal explanations, which is beyond the scope of the methodology and study. As (Batty 2012) summarizes this succinctly, '[models to quantify cities] are being used increasingly to 'inform' rather than 'predict' as a new relativism sweeps the field'.

Contributions and Conclusion

The World Economic Forum estimates that 56.2% of the world's population lives in cities where North Americans are leading this urbanization trend, with 83.6% of its residents living in cities in 2020. At a theoretical level, this paper presents a methodology that combines social science thinking with data science principles to design a system to pulse cities as they continue to evolve through time. By collecting a 2-year panel data, we show that this design methodology is theoretically rigorous and practically feasible in highly urbanized environments where data is ubiquitous. At a practical level, this project is the first step to establishing empirical evidence that big data within a city can be used to alert city planners of emerging red flags. The alert notification system will help government and other policymakers make the best use of the wealth of data we have today in a more meaningful, systematic, and theoretically driven fashion. In the follow-up work, we will outlier specific cases of outliers within the city and validate these outliers with additional data collection, such as using interviews and surveys to illustrate the possibility of a programmatic approach toward measuring and validating city phenomena. This work presents one contribution that information systems research should have to societies at large.

Acknowledgments

This project is supported by the National Research Foundation, Singapore, and Ministry of National Development, Singapore under its Cities of Tomorrow R&D Programme (Award No: COT-CityScan-2020-1). Any opinions, findings and conclusions or recommendations expressed in this material are those of the author(s) and do not reflect the views of National Research Foundation, Singapore and Ministry of National Development, Singapore

References

Anderson, E. 2004. "The Cosmopolitan Canopy," *The Annals of the American Academy of Political and Social Science* (595:1), pp. 14-31.

Barnaghi, P., Tönjes, R., Höller, J., Hauswirth, M., Sheth, A., and Anantharam, P. 2014. "Citypulse: Real-Time Iot Stream Processing and Large-Scale Data Analytics for Smart City Applications," *Europen semantic web conference (ESWC)*: sn.

Batty, M. 2008. "The Size, Scale, and Shape of Cities," *science* (319:5864), pp. 769-771.

Batty, M. 2012. "Building a Science of Cities," *Cities* (29), pp. S9-S16.

Batty, M. 2013. *The New Science of Cities*. MIT press.

Blumer, H. 1969. *Symbolic Interactionism*. Englewood Cliffs, NJ: Prentice-Hall.

Collins, R. 1988. "The Micro Contribution to Macro Sociology," *Sociological theory* (6:2), pp. 242-253.

Froehlich, J. E., Neumann, J., and Oliver, N. 2009. "Sensing and Predicting the Pulse of the City through Shared Bicycling," *Twenty-First International Joint Conference on Artificial Intelligence*.

Gabaix, X. 1999. "Zipf's Law for Cities: An Explanation," *The Quarterly journal of economics* (114:3), pp. 739-767.

Giatsoglou, M., Chatzakou, D., Gkatziaki, V., Vakali, A., and Anthopoulos, L. 2016. "Citypulse: A Platform Prototype for Smart City Social Data Mining," *Journal of the Knowledge Economy* (7:2), pp. 344-372.

Glaeser, E. 2012. *Triumph of the City: How Our Greatest Invention Makes Us Richer, Smarter, Greener, Healthier, and Happier*. Penguin.

Hasija, S., Shen, Z.-J. M., and Teo, C.-P. 2020. "Smart City Operations: Modeling Challenges and Opportunities," *Manufacturing & Service Operations Management* (22:1), pp. 203-213.

Kose, E., Vural, D., and Canbulut, G. 2020. "The Most Livable City Selection in Turkey with the Grey Relational Analysis," *Grey Systems: Theory and Application*).

McKenzie, G., Janowicz, K., Gao, S., Yang, J.-A., and Hu, Y. 2015. "Poi Pulse: A Multi-Granular, Semantic Signature–Based Information Observatory for the Interactive Visualization of Big Geosocial Data," *Cartographica: The International Journal for Geographic Information and Geovisualization* (50:2), pp. 71-85.

Miranda, F., Doraiswamy, H., Lage, M., Zhao, K., Gonçalves, B., Wilson, L., Hsieh, M., and Silva, C. T. 2016. "Urban Pulse: Capturing the Rhythm of Cities," *IEEE transactions on visualization and computer graphics* (23:1), pp. 791-800.

Obaid, M., Kurdyukova, E., and Andre, E. 2012. "City Pulse: Supporting Going-out Activities with a Context-Aware Urban Display," *International Conference on Advances in Computer Entertainment Technology*: Springer, pp. 529-532.

Puiu, D., Barnaghi, P., Tönjes, R., Kümper, D., Ali, M. I., Mileo, A., Parreira, J. X., Fischer, M., Kolozali, S., and Farajidavar, N. 2016. "Citypulse: Large Scale Data Analytics Framework for Smart Cities," *IEEE Access* (4), pp. 1086-1108.

Rozenfeld, H. D., Rybski, D., Andrade, J. S., Batty, M., Stanley, H. E., and Makse, H. A. 2008. "Laws of Population Growth," *Proceedings of the National Academy of Sciences* (105:48), pp. 18702-18707.

Snow, D. A. 2001. "Extending and Broadening Blumer's Conceptualization of Symbolic Interactionism," *Symbolic interaction* (24:3), pp. 367-377.

Snow, D. A. 2013. "Framing and Social Movements," *The Wiley-Blackwell Encyclopedia of Social and Political Movements*).

Turner, J. H. 2011. "Extending the Symbolic Interactionist Theory of Interaction Processes: A Conceptual Outline," *Symbolic Interaction* (34:3), pp. 330-339.

Vaccari, A., Martino, M., Rojas, F., and Ratti, C. 2010. "Pulse of the City: Visualizing Urban Dynamics of Special Events,").