Big data kills the virus

Computational Socioeconomics: A data-driven framework for quantifying progress towards achieving the Sustainable Development Goals (SDGs)

Jian Gao, Tao Zhou, Quanhui Liu
Insitution of New Economic Development

The improvements of data acquisition and processing capabilities, as well as artificial intelligence and statistical mechanics, have rapidly and significantly changed the methodology of social and economic research. The recent paradigm shifting of social science driven by big data and artificial intelligence provides promising and novel data-driven methods for measuring the progress of Sustainable Development Goals (SDGs). This shift affects areas ranging from no poverty to good health and well-being, from gender equality to quality education, and from economic growth to innovation and infrastructure. Governments at both national and regional levels can benefit from leveraging new methods under the framework of Computational Socioeconomics[1] to better assess their progress towards sustainable development over space and time with a higher efficiency and a lower cost.

New Methodology Shifts

Social and economic studies become increasingly dependent on real data. Yet, the traditional way to obtain real data has many limitations. For example, larger-scale and more precise data usually consumes huge resources and lacks timeliness. Fortunately, thanks to the digital wave that swept across the world in recent decades, social and economic researchers face an unprecedented opportunity to develop a quantitative methodology.

Data in the processes of socioeconomic development and human activities are recorded by an increasing number of sensing devices, online platforms, and other data acquisition terminals such as remote-sensing satellites, mobile phones, social media platforms, and online trading platforms.[2, 3] On the other hand, these data are of larger size, almost in real time and with higher resolution, can reduce the sparsity and bias in small-size data as well as reduce the invisible parts in the developing processes. Therefore, based on these large-scale novel data, we can in principle make great progress in perceiving socioeconomic situations, evaluating development progresses, predicting future social and economic trends, and so on.[4]

The increasing volume and diversity of novel data lead to methodological changes in two aspects. Firstly, simple statistical tools are not suitable for analyzing unstructured data such

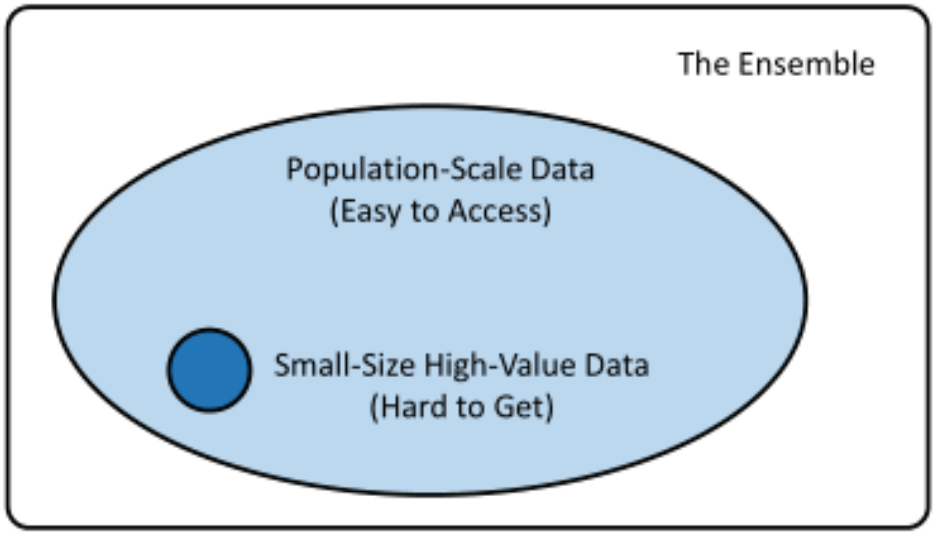


Figure 1. Illustration of ensembling novel data in the computational socioeconomics framework. © **Gao, Zhang & Zhou. Physics Reports, 2019**

as street view images and textual contents. Researchers are in serious need of more advanced techniques of data mining and machine learning.[5] Secondly, with population-scale data, one can concentrate on a small sampled subset and add high-valued new dimensions of data.

New data dimensions can be obtained through traditional means such as a questionnaire survey. A model can be trained based on the small sample to infer new dimensions from the original ones. After applying the model to the whole dataset, one can obtain new data dimensions for all individuals.

This method integrates some routine methods like sampling, labeling, and surveying, while it is more powerful in practice. For example, it is relatively easy to obtain the population-scale data on mobile communication and mobility, while it is very hard to know the household income of every family without compiling a population-scale economic census. Under the new framework, we first obtain household incomes of some families via routine questionnaires. Then, using the small dataset, we train machine learning models to predict household income of a family based on the mobile phone data of the family members.

Although the inferred data is not perfect, it can be very close to the real data under a certain well-designed algorithm. Notice, a significant advantage is that the high-value data for almost every individual can be obtained at a very low cost. Combining the accessible population-scale data, a small sample of high-value but hard-to-get data, and a properly selected or well-designed algorithm to infer the high-value data for individuals other than the sample is a novel and representative method in the computational socioeconomics study (Figure 1), showing the deep integration of social science and computer science methods.

Nowcasting Poverty and Growth

Revealing the status of social and economic development in a near real-time manner and with a lower cost is one of the long-standing problems that hinders the effects towards Sustainable Development Goals (SDGs). To approach the goals of no poverty, the first step is to accurately map the spatial distribution of poverty. New data and tools introduced in computational socioeconomics have been utilized to better reveal, explain and predict global poverty and economic growth, such as data from remote sensing (RS) and mobile phone (MP).

High resolution data from RS, for example, nighttime lights (NTLs) satellite imagery, has been used to supply information about economic activity, especially in developing countries where traditional economic census data are insufficient. NTLs data can provide an unambiguous indication of the spatial distribution of economic development. For example, Jean et al.[6] applied deep learning algorithms to learn the relationship between NTLs and daytime satellite imagery. The former can predict the wealth distribution while the latter contains rich information about landscape features. The image features extracted from the daytime imagery can explain up to 75% of the variation in the average household asset across five African countries. Moreover, the method is able to reconstruct survey-based indicators of regional poverty with high accuracy.

MPs are able to capture an enormous information and provide cost-effective data at the individual level. With MP logs related to consumption and expenditure, socioeconomic status can be inferred by employing machine learning approaches at the aggregated subnational and national levels. For example, Blumenstock et al.[7] presented a novel method to explore the relationship between MP usages and wealth in developing countries. By analyzing the data from Rwanda, they found that household expenditures are positively correlated with MP usages, for instance, in the number of different districts contacted. Moreover, by applying a machine learning approach to analyze the follow-up phone surveys of some individual subscribers, Blumenstock et al.[8] showed that individual wealth can be well predicted and individuals in relative poverty can be accurately identified. Then, they generated out-of-sample predictions for 1.5 million MP users and produced the wealth map of Rwanda at a very high resolution and accuracy. This method is promising to map the distribution of wealth and other socioeconomic indicators for the full national population.

Understanding how economies develop to prosperity is a long-standing challenge in economic growth. Hidalgo and Hausmann [9] proposed a novel index named economic complexity (ECI), a non-monetary metric which quantitatively assesses a country's potential for future economic growth. In particular, a Method of Reflections (MR) is proposed to characterize the structure of ``country-product'' bipartite network in international trade and the variables produced by the MR method can be interpreted as indicators of economic complexity. Empirical results showed that countries' ECIs are highly correlated with their income levels and are predictive of their

Moreover, individual behaviors on social networking platforms have been used to estimate individual personality and mental states such as depression and suicidal intent.

future growth. Later, a statistical approach is employed to define a new set of metrics and to quantify the fitness of countries and the complexity of products. Tacchella et al.[10] showed that this scheme outperforms the International Monetary Fund (IMF) five-year GDP per capita forecast by more than 25% in accuracy, and the method's forecasting errors are predictable. These complexity and fitness measures have been used to quantify the economic complexity and development at different spatial resolutions, such as China's regional economic complexity.[11]

Perception of Regions and Cities

High-resolution data and improved methods allow us to reveal economic activity and socioeconomic status in subnational, regional, and urban scales. For example, indicators derived from both nighttime lights (NTLs) and very high resolution (VHR) imagery have been used to map poverty at fine scales. In particular, novel data from mobile phone (MP) and Google Street Views provide a promising way to the perception of cities and communities.

Slums are common in low- and middle-income countries with poor quality of basic services (e.g., water supply, electricity, and sanitation). Detecting and monitoring slum areas is valuable for implementing policies to improve living conditions. Recently, VHR images have been increasingly used to inventory the location and physical composition of slums. For example, Kit et al.[12] developed the concept of lacunarity to identify slums in Hyderabad,

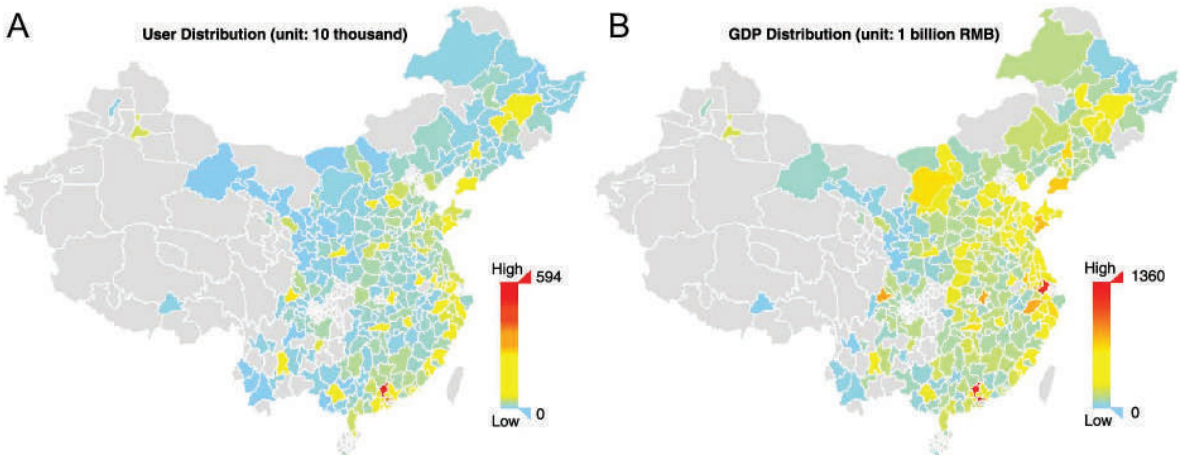


Figure 2. The spatial distributions of the intensities of online activity (A) and the values of GDP (B) in prefecture-level cities of China in 2012. © **Liu et al. Physica A, 2016**

India. The best method can reach an accuracy of 0.8333 in slum identification and can capture the changing patterns of slum areas from 2003 to 2010. Similarly, Kuffer et al.[13] utilized the gray-level co-occurrence matrix (GLCM) variance to distinguish slum areas in VHR imagery and showed that the overall accuracy can be increased to 90% by adding spectral information to the GLCM within a random forest classifier.

Social media (SM) data have been used to track socioeconomic well-beings. For example, based on the registered location information of nearly 200 million Weibo users in China, Liu et al. [14] explored the relationship between online activities and socioeconomic indices (Figure 2). They found that UN is strongly correlated with socioeconomic indices, suggesting that socioeconomic status can be inferred from online social activity at the city-level. Of particular significance, they further proposed a method to detect a few abnormal cities, whose GDP is much higher than others with the same number of registered users. Similarly, with data of friendship information and geo-locations from Gowalla in the US, Holzbauer et al.[15] studied the relations between regional economic status and quantitative measures of social ties. They found that cross-state long ties are strongly correlated with three economic measurements, namely, GDP, the number of patents, and the number of startups.

Crowdsourcing methods and computational vision techniques have been used to measure livability, safety and inequality, to infer the status of urban life, and to quantify the changes of urban streetscapes. For example, Salesses et al.[16] presented a method to measure the urban perception of safety, class, and uniqueness in two US cities and two Austrian cities based on hundreds of geotagged images. They found that the two US cities are perceptually more unequal, and that the spatial variation of urban perception helps explain violent crimes in NYC zones

at zip-code resolution. Later, Naik et al.[17] trained a scene understanding model named Streetscore based on data from an online survey to predict the perceived safety of a streetscape using generic image features. Physical appearances of neighborhoods are not static but changing over time. Naik et al.[18] introduced a computer vision method to understand physical dynamics of cities based on street views at different times. They found that education and population density, physical proximity to city centers, and better initial appearances are associated with physical improvements in neighborhoods.

Deep-learning-based computer vision techniques have been applied to analyze digital imagery, which provides a faster and cheaper alternative of community survey. For example, Gebru et al.[19] proposed a method to estimate socioeconomic trends from 50 million street view images in 200 US cities. They automatically detected 22 million distinct vehicles from images using the object recognition algorithm and then deployed CNNs to determine features of vehicles and classify each vehicle into one of the 2,657 fine-grained categories. Using the resulting data, they estimated race and education levels by training a logistic regression model and estimated income and voter preferences by employing a ridge regression model. Compared to the American Community Survey, their demographic estimates exhibit satisfied accuracy at the city level. The method can also provide a good accuracy at a more fine-grained zip code resolution; for example, the estimation of the percentage of Asians yields a high correlation at zip code resolution for Seattle.

Gender Equality and Social Segregation

Demographic attributes of individuals have remarkable effects on their socioeconomic status, while traditional methods of individual profiling based on surveys and censuses are costly and follow a long-time delay. Recently, data from novel sources such as social media (SM) and mobile phones (MPs) have been used alternatively to predict individual demographic attributes and to analysis social and religious segregations. Moreover, individual behaviors on social networking platforms have been used to estimate individual personality and mental states such as depression and suicidal intent.

MP and online data have been used to infer demographic information – gender in particular. Frias-Martinez et al.[20] analyzed call detail records (CDRs) and found that male and female users are significantly different in behavioral and social variables such as duration of calls and degree in social networks. They proposed a semi-supervised classification algorithm that can identify gender with an accuracy up to 0.80. Felbo et al.[21] developed a convolutional network architecture to transform MP data into high-level features for each week and then aggregated patterns across weeks by reusing the same convolutional filters. They designed a 2-step model using an SVM with a radial basis function kernel, which slightly outperforms the state-of-the-art method, with an accuracy 0.797 in gender prediction. On the exposure of online platforms to different genders, Mislove et al.[22] inferred gender of Twitter users representing more than 1% of

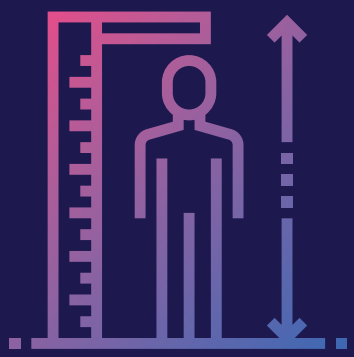
the U.S. population based on their first names and found that 71.8% of the users had a male name, showing a strong gender bias of Twitter towards male users. On the height premium in labor market, Yang et al.[23] found stronger effects of height premium on female than on male after analyzing a dataset covering over 140,000 Chinese job seekers. Of particular, they found that the gender differences decrease as the education level increases and become insignificant after holding all control variables fixed.

By leveraging novel large-scale data, urban segregation of people with different socioeconomic status have been studied. For example, Shelton et al.[24] developed an approach to study intra-neighborhood segregation, mobility and inequality based on geotagged tweets in Louisville. They proposed to understand Louisvillian neighborhoods by the fluid, porous, and actively produced. Similarly, Yip et al.[25] analyzed the mobility patterns of people in Hong Kong that are tracked by a mobile phone app. They found that the interactions of people with other income groups are limited. Rich people tend to move to rich neighborhoods, while poorer people tend to move to poorer neighborhoods. Recently, Louf and Barthelemy[26] provided a direct definition of residential segregation and showed that the richer class in high density zones is overrepresented. In particular, they suggested density as a relevant factor for understanding urban income structure and explaining differences observed in cities.

Data from social networks have been used to study religious segregation and urban indigenization. Hu et al.[27] quantified religious segregation by analyzing religious social network based on Weibo. They found that the religious network is highly segregated, and the extent of religious segregation is higher than racial segregation. In addition, 46.7% of cross-religion connections are probably related to charitable issues, suggesting the role of charitable activities in promoting cross-religion communications. Yang et al.[28] identified the distinct mobility patterns of natives and non-natives in five large cities in China by analyzing about 1.37 million check-ins. They found that the distribution of location visiting frequencies is relatively homogeneous for natives as they usually check in repeatedly at locations of personal importance. By contrast, the distribution is more heterogeneous for non-natives as they tend to visit popular locations. With this insight, Yang et al.[28] proposed a so-called indigenization coefficient to estimate the likelihood of an individual to be a native or to what extent an individual behaves like a native, which is based solely on check-in behaviors. Such method can be applied in estimating the time required for non-natives to behave the same as natives as well as in enhancing the prediction accuracy of human mobility.

Climate Action and Disaster Relief

Climate change and disaster surveillance is critical to social and economic systems. Along with increased urbanization and changing climate, many areas are now facing an unprecedented number of emergent events and natural disasters, which pose numerous threats to human life and economic development.



On the height premium in labor market, Yang et al found stronger effects of height premium on female than on male after analyzing a dataset covering over 140,000 Chinese job seekers

It urges rapid situational awareness and efficient management strategies to reduce human suffering and economic losses. In rural areas, assessments of natural hazards usually follow a delay, resulting in difficulties of disaster response and relief. In urban areas, detections of natural disasters (such as earthquakes, floods and hurricanes) are critical not only for governments' rapid disaster response but also for in-depth understanding of human behaviors in extreme situations that will help in better designing strategies in disaster relief.

Novel data sources have been leveraged to improve emergency awareness and disaster management such as remote sensing (RS), mobile phone (MP), and social media (SM), with remarkable advantages of low acquisition cost, real-time updates and high spatio-temporal resolutions. In particular, deep learning algorithms have been introduced to analyze RS data for rapid earthquake damage mapping. For the 2010 Haiti Earthquake, Cooner et al.[29] evaluated the effectiveness of several deep learning algorithms in detecting earthquake damage. They found that spatial texture and structure features extracted from satellite images can detect damaged buildings with an error rate below 40% under a multilayer feedforward neural network framework. Similarly, Bai et al.[30] developed a deep learning algorithm to map damage due to the 2011 Tohoku Earthquake-Tsunami. Their algorithm can classify damage with an overall accuracy 0.709 based on pre- and post-disaster images.

Rapid emergency detection based on mobile phone (MP) data can facilitate humanitarian response and reduce the toll of extreme events. Based on the combined data of MP activities and official event records in Rwanda, Dobra et al.[31] proposed an efficient system that can detect days with anomalous behavioral patterns under many emergent and non-emergent events. MP data have also been used to assess population displacements and improve emergency responses during large-scale disasters. For example, Lu et

al.[32] explored the predictability of population displacements after the Haiti earthquake. They found that the population in PaP decreases by 23% in the three-month period after the earthquake due to population movements. Also, the destinations of people who left PaP during the first three weeks correlate well with their mobility patterns during normal times.

Social media (SM) is a valuable source of information for gaining situational awareness, detecting and locating emergent events, improving disaster response, and enhancing relief efforts. Indeed, the utilization of SM data has transformed the methodology of earthquake detection and early warning[33], where the distribution of shakings can be mapped in minutes from earthquake-related posts. For example, Acar et al.[34] studied earthquake information sharing on Twitter by analyzing the tweets posted near two disaster-struck areas during the 2011 Tohoku Earthquake. They found that people in directly affected areas tweeted to announce their uncertain and unsafe situation, while people in remote areas tweeted to inform followers that they are safe. SM data have been increasingly used in monitoring and mapping floods in a timely manner. For example, Arthur et al.[35] leveraged tweets to detect and locate flood events in the UK. They collected tweets containing flood-related terms and located flood events by analyzing many indicators such as mentioned place names and GPS coordinates. They produced high-quality flood event maps based on the relevant geotagged tweets and validated the flood maps by official data.

SDG3: Good Health and Well-being

Ensuring the healthy lives and promoting well-being for all are the goals of the SDG3. With the coming of big data and the development economic, the last decade has made the significant strides in increasing the life expectancy and reducing the infant and maternal mortality rates. With the availability of various data sources, lots of data-driven model have been developed and major progresses have also made in preventing the spread of the communicable diseases, such as malaria, the seasonal influenza, and the pandemic, and so on.

A report given by the Centers for Disease Control and Prevention (CDC) shows that, an average of 28.41 million cases, 461 111 hospitalizations, and 40 500 influenza related deaths occur in each year from 2005 to 2018 in the United States,[36] which caused the economic burden at $5.8 billion annually.[37] As the globular head of hemagglutinin is evolving continually, the efficacy of the seasonal vaccines depends on the match between the antigens included in the vaccine and those presented by circulating influenza strains. Sah et al[38] assumed that if 10% of typical seasonal vaccines is replaced with 75% efficacious universal vaccine, according to the data-driven model accounted for various data monitored by the CDC, they showed that about 5.3 million cases, 81 000 hospitalizations, and 6300 influenza-related deaths per year would be averted. With the availability of the weekly temperature, relative

© Shutterstock

humidity and atmospheric pressure data for each city of 603 cities in the United States, Dalziel et al[46] developed a climate-forced susceptible-exposed-infected-removed-susceptible compartmental model for influenza epidemics. They find that the period of the influenza season in the smaller cities is shorter, and the city-level incidence data is positively correlated with population size , and further their study reveal that the urban centers incubate critical chains of transmission outside of peak climatic conditions, altering the spatiotemporal geometry of herd immunity. For the understanding and prediction of the epidemic, Liu et al.[39] built a subset of the Italian and Dutch populations with the highly detailed sociodemographic data. By calibrating the epidemic model with the empirical epidemiological data, they show that the classical concept of the basic reproduction number is untenable in realistic populations. Litvinova et al.[40] performed a diary-based contact survey estimating the patterns of social interactions before and during the implementations of reactive school-closure strategies in the influenza season, and it is incorporated the macro sociodemographic data. With this innovative hybrid survey-modeling framework, they showed that the gradual reactive school-closure policies can mitigate the spread of influenza.

The emergence of the innovative infectious diseases, such as the SARS epidemic of 2003, the 2009 H1N1 influenza, and most recently the 2019_nCoV, affects the lives of tens of thousands or even millions of people. As the absence of the vaccine for the emergent infectious diseases and the globalization, the highly virulent innovative diseases increase the risk of every city in the world being invaded. Brockmann et al.[41] based on the air-traffic data defined the effective distance which predicts the disease arrival times of the invaded city accurately. Their method also works well for both the worldwide 2009 H1N1 influenza pandemic and 2003 SARS epidemic. Zhang et al.[42] developed a data-driven global stochastic epidemic model, accounted for the real-world demographic, human mobility, socioeconomic, temperature, and the vector density data, for the spread of the Zika virus (ZIKV) in the Americas. They estimate the time of first introduction of ZIKV to Brazil, and also revealed that the spreading features of ZIKV. For the new coronavirus originated in Wuhan, China, Chinazzi et al.[43] developed a detailed individual based mobility model which covers more than 3300 subpopulations in about 190 countries/territories. By using the cases detected outside China, they estimate the potential outbreak size in Wuhan and the basic production number. Similar to Chinazzi's report, by using the cases detected in overseas, Imai et al.[44] accorded to the traffic of Wuhan Interantion Airporte and estimate the basic reproduction number close to Chinazzi's. [43]

Visions and Actions

The availability of large-scale and high-resolution data from social and economic systems has provides a new way to improve urban spatial equity. For example, Louail et al. [45] analyzed a database of card transactions

in two Spanish cities and then proposed a bottom-up approach to redistribute money flows for equality situations through redirecting a limited fraction of individual shopping trips. They constructed the “individual-business” bipartite spatial network, where the edges correspond to card transactions. Then, they performed the rewiring of individual transactions by redirecting them to the same business category located in different neighborhoods. The goal was to re-balance the commercial income among neighborhoods and with the preservation of human mobility properties. They found that reassigning only 5% of individual transactions can reduce more than 80% spatial inequality between neighborhoods and can even improve other sustainability indicators like total distance traveled and spatial mixing. Their work illustrates an excellent implementation of crowdsourcing; the “Robin Hood effect”, a process through which capital is redistributed to reduce inequality.

Methods and data sources introduced in the Computational Socioeconomics can benefits the actions towards achieving SDGs and the evaluation of the progress. Specially, the above-mentioned novel perspective and methodology, driven by big data and artificial intelligence, will promisingly become the mainstream research framework in the action of SDGs.

1. Gao, J., Y.-C. Zhang, and T. Zhou, Computational socioeconomics. Physics Reports, 2019. 817: p. 1-104.

2. Gao, J. and T. Zhou, Big data reveal the status of economic development. Journal of University of Electronic Science and Technology of China, 2016. 45(4): p. 625-633.

3. Mayer-Schonberger, V. and K. Cukier, Big Data: A Revolution That Will Transform How We Live, Work, and Think. 2013, New York, NY, USA: Houghton Mifflin Harcourt.

4. Gao, J., Research on the Spatial Structure and Dynamics of Socio-Economic Systems. 2019, University of Electronic Science and Technology of China.

5. Lecun, Y., Y. Bengio, and G. Hinton, Deep learning. Nature, 2015. 521(7553): p. 436-444.

6. Jean, N., et al., Combining satellite imagery and machine learning to predict poverty. Science, 2016. 353(6301): p. 790-794.

7. Blumenstock, J., Y. Shen, and N. Eagle. A method for estimating the relationship between phone use and wealth. in QualMeetsQuant Workshop at the 4th ACM/IEEE International Conference on Information and Communication Technologies and Development. 2010. ACM Press.

8. Blumenstock, J., G. Cadamuro, and R. On, Predicting poverty and wealth from mobile phone metadata. Science, 2015. 350(6264): p. 1073-1076.

9. Hidalgo, C.A. and R. Hausmann, The building blocks of economic complexity. Proceedings of the National Academy of Sciences of the United States of America, 2009. 106(26): p. 10570-10575.

10. Tacchella, A., D. Mazzilli, and L. Pietronero, A dynamical systems approach to gross domestic product forecasting. Nature Physics, 2018. 14(8): p. 861-865.

11. Gao, J. and T. Zhou, Quantifying China’s regional economic complexity. Physica A: Statistical Mechanics and its Applications, 2018. 492: p. 1591-1603.

12. Kit, O., M. Lüdeke, and D. Reckien, Texture-based identification of urban slums in Hyderabad, India using remote sensing data. Applied Geography, 2012. 32(2): p. 660-667.

13. Kuffer, M., et al., Extraction of slum areas from VHR imagery using GLCM variance. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2016. 9(5): p. 1830-1840.

14. Liu, J.-H., et al., Online social activity reflects economic status. Physica A: Statistical Mechanics and its Applications, 2016. 457: p. 581-589.

15. Holzbauer, B.O., et al. Social ties as predictors of economic development. in Proceedings of the 12th International Conference and School on Advances in Network Science. 2016. Springer.

16. Salesses, P., K. Schechtner, and C.A. Hidalgo, The collaborative image of the city: Mapping the inequality of urban perception. PLoS ONE, 2013. 8(7): p. e68400-e68400.

17. Naik, N., et al. Streetscore--Predicting the perceived safety of one million streetscapes. in 2014 IEEE Conference on Computer Vision and Pattern Recognition Workshops. 2014. IEEE Press.

18. Naik, N., et al., Computer vision uncovers predictors of physical urban change. Proceedings of the National Academy of Sciences of the United States of America, 2017. 114(29): p. 7571-7576.

19. Gebru, T., et al., Using deep learning and Google Street View to estimate the demographic makeup of neighborhoods across the United States. Proceedings of the National Academy of Sciences of the United States of America, 2017. 114(50): p. 13108-13113.

20. Frias-Martinez, V., E. Frias-Martinez, and N. Oliver. A gender-centric analysis of calling behavior in a developing economy using call detail records. in Proceedings of the 2010 AAAI Spring Symposium: Artificial Intelligence for Development. 2010. AAAI Press.

21. Felbo, B., et al. Modeling the temporal nature of human behavior for demographics prediction. in Proceedings of the European Conference on Machine Learning and Knowledge Discovery in Databases. 2017. Springer

22. Mislove, A., et al. Understanding the demographics of Twitter users. in Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media. 2011. AAAI Press.

23. Yang, X., et al., Height conditions salary expectations: Evidence from large-scale data in China. Physica A: Statistical Mechanics and its Applications, 2018. 501: p. 86-97.

24. Shelton, T., A. Poorthuis, and M. Zook, Social media and the city: Rethinking urban socio-spatial inequality using user-generated geographic information. Landscape and Urban Planning, 2015. 142: p. 198-211.

25. Yip, N.M., R. Forrest, and X. Shi, Exploring segregation and mobilities: Application of an activity tracking app on mobile phone. Cities, 2016. 59: p. 156-163.

26. Louf, R. and M. Barthélemy, Patterns of residential segregation. PLoS ONE, 2016. 11(6): p. e0157476-e0157476.

27. Hu, J., Q.-M. Zhang, and T. Zhou, Segregation in religion networks. EPJ Data Science, 2019. 8: p. 6-6.

28. Yang, Z., et al., Indigenization of urban mobility. Physica A: Statistical Mechanics and its Applications, 2017. 469: p. 232-243.

29. Cooner, A.J., Y. Shao, and J.B. Campbell, Detection of urban damage using remote sensing and machine learning algorithms: Revisiting the 2010 Haiti Earthquake. Remote Sensing, 2016. 8(10): p. 868-868.

30. Bai, Y., E. Mas, and S. Koshimura, Towards operational satellite-based damage-mapping using U-Net Convolutional Network: A case study of 2011 Tohoku Earthquake-Tsunami. Remote Sensing, 2018. 10(10): p. 1626-1626.

31. Dobra, A., N.E. Williams, and N. Eagle, Spatiotemporal detection of unusual human population behavior using mobile phone data. PLoS ONE, 2015. 10(3): p. e0120449-e0120449.

32. Lu, X., L. Bengtsson, and P. Holme, Predictability of population displacement after the 2010 Haiti earthquake. Proceedings of the National Academy of Sciences of the United States of America, 2012. 109(29): p. 11576-11581.

33. Allen, R.M., Transforming earthquake detection? Science, 2012. 335(6066): p. 297-298.

34. Acar, A. and Y. Muraki, Twitter for crisis communication: Lessons learned from Japan's tsunami disaster. International Journal of Web Based Communities, 2011. 7(3): p. 392-402.

35. Arthur, R., et al., Social sensing of floods in the UK. PLoS ONE, 2018. 13(1): p. e0189327-e0189327.

36. Centers for Disease Control and Prevention, Seasonal influenza vaccine effectiveness, 2005–2018. https://www.cdc.gov/flu/professionals/vaccination/effectiveness-studies.htm. Accessed 26 July 2018.

37. Ozawa S., et al., Modeling the economic burden of adult vaccine-preventable diseases in the United States. Health Aff. (Millwood), 2016.35, 2124–2132.

38. Sah P., et al. Future epidemiological and economic impacts of universal influenza vaccines. PNAS, 2019.116(41), 20786-20792.

39. Liu Q.-H., et al. Measurability of the epidemic reproduction number in data-driven contact networks. PNAS , 2018.115(50), 12680-12685.

40. Litvinova M., et al. Reactive school closure weakens the network of social interactions and reduces the spread of influenza. PNAS, 2019.116(27), 13174-13181.

41. Brockmann D., et al. The Hidden Geometry of Complex, Network-Driven Contagion Phenomena. Science, 2013. 342, 1337-1342.

42. Zhang Q., et al. Spread of Zika virus in the Americas. PNAS, 2017.114, E4334-E4343.

43. Chinazzi M., et al. Series Reports Entitled "Preliminary assessment of the International Spreading Risk Associated with the 2019 novel Coronavirus (2019-nCoV) outbreak in Wuhan City" (unpublished).

44. Imai N., et al. Transmissibility of 2019-nCoV (unpublished).

45. Louail, T., et al., Crowdsourcing the Robin Hood effect in cities. Applied Network Science, 2017. 2: p. 11-11.