

The new data landscape for regional and urban analysis

Xinyue Ye · Canfei He

Published online: 12 August 2016
© Springer Science+Business Media Dordrecht 2016

Introduction

The world's population has been gradually concentrated in urban areas at an extraordinary scale and speed. Hence, regional and urban analysis faces many challenges as respective contexts are experiencing rapid change. For example, people's livelihoods in the metropolitan areas have been severely influenced by the increasingly frequent natural or man-made disasters (Wang et al. 2016). With the proliferation of social media and the now ubiquitous smart phones, text messages have been generated and spread in large volumes every day (Ye et al. 2016b). These changing regional and urban contexts demand innovative spatial thinking that will capture the patterns and processes, providing spatial strategies for sustainable development. The research agenda is being substantially transformed and redefined in light of the new data, especially big data. An era of big data that are characterized by high volume, velocity, variety, exhaustivity, resolution and indexicality, as well as

relationality and flexibility is approaching (Laney 2001; Kitchin 2013). The size of big data is often beyond the ability of typical database software tools to capture, store, manage and analyze, but the word "big" is rather relative (Manyika et al. 2011). What is more important is the depth and breadth of the data that allows potentially unprecedented insights into our world, and thus transforms the ways in which we can sense and study the world (Gonzalez-Bailon 2013). This special issue plans to focus on the development of theories, methods and practices on this collaborative and interdisciplinary frontier for this new reality. This special section contains eight articles and rejected four manuscripts through a rigorous peer-review process.

The amount and diversity of new data sources relating to regions and cities has grown dramatically in complex ways at degrees of detail and scope unthinkable. Such data is big, spatial, temporal, dynamic, and unstructured. This trend has fostered a growing research community aiming at maximizing the potential of massive data to improve human well-being (Ye et al. 2016a). In the special issue, "Revealing the relationship between spatiotemporal distribution of population and urban function with social media data" utilized Tencent, one of the biggest Internet companies in China, to measure urban morphologies during various time periods. The result shows that as urban functions become more mixed, the temporal distribution of the population tends to be more stable. "Research on China's city network based on users' friend relationships in online social networks: a case

X. Ye (✉)
Department of Geography and Computational Social
Science Lab, Kent State University, Kent, OH 44242,
USA
e-mail: xye5@kent.edu

C. He
College of Urban and Environmental Sciences, Peking
University, Beijing 100871, China
e-mail: hecanfei@pku.edu.cn

study of Sina Weibo” examines China’s urban network based on Sina Weibo, the most popular online social networking service in China. This paper argues that the role of a city in the real world has significant influence on its power in the city network through online social networking connections. “C-IMAGE: City Cognitive Mapping Through Geo-Tagged Photos” explores the interactions between city and human perception based on a number of photos from 26 different cities through the metadata and image content. This paper verifies Kevin Lynch’s five elements of urban image: node, path, edge, district, and landmark.

Regional and urban analysis is shifting towards analyzing ever-increasing amounts of large-scale, diverse data in an interdisciplinary, collaborative and timely manner. Insight into the new data landscape will not only test existing urban and regional development theories, but also contribute to problem solving in the globalizing world (Ye 2016). Much of the big data such as transactional data, transportation monitoring data and social media data are geographically and temporally referenced and create remarkable opportunities for social scientists, particularly human geographers. With the emergence of new data collection technologies, advanced data mining and analytics support, big data driven research that can leverage micro-level, meso-level and macro-level suggests the possibility of a scientific paradigm shift toward computational social science (Chang et al. 2014).

The availability of large-size geo-referenced datasets along with the high performance computing technology has raised fundamental challenges and opportunities to mainstream social science research on whether and how these new trends in data and technology can be utilized to help detect the trends and patterns in the urban and regional dynamics. Analyzing large-scale data requires an innovative design of algorithms and computing resources. Big data analytics gives birth to big opportunities for social scientists and decision-makers due to its great power on visualization, prediction and simulation, transforming the way we can approximate the world. Big data analytics also provides broader and deeper information for behavioral analysis, trend analysis, spatial analysis and network analysis. To some extent, big data offer a panoramic view of correlation. Despite these potential benefits generated by the introduction of big data into our research, we

should be cautious about how to embrace big data and how big data might be integrated into preexisting structures of scholarly knowledge production (Graham and Shelton 2013). Traditional social science research relies heavily on the extraction of information from small number of observations, for example, through interviews, surveys and a handful of case studies. In contrast, big data can track what we do, the time and place of our actions and the chains of interdependence that link those actions together, helping us draw a richer, more detailed, timely and interrelated picture of regional and urban dynamics (Gonzalez-Bailon 2013). With big data, social scientists can build better models and more dynamic maps of how people interact with places, how those places are perceived and how they come to be. Big data thus holds the promise of moving from data-scarce to data-rich, from static to dynamic, from relatively simple to more complex and sophisticated regional and urban studies (Kitchin 2013).

In this special issue, “Emerging Data Sources and the Study of Genocide: A Preliminary Analysis of Prison Data from S-21 Security-Center, Cambodia” synthesizes diverse sources of information to provide more robust analyses of the patterns and trends of mass violence. This article develops a database using information from a security-center (S-21) associated with the Cambodian genocide (1975–1979). “A micro-level analysis of firearm arrests’ effects on gun violence in Houston, Texas” explores the citywide space–time interaction between shootings and firearm arrests in Houston, Texas. This paper also investigates the deterrence or escalation effects due to arrests. “Implementing a real-time Twitter-based system for resource dispatch in disaster management” aims to discover and utilize relevant tweets in disaster management. The study develops a Web GIS platform for geo-tagged tweets operation. “Locating healthcare facilities using a network-based covering location problem” deals with healthcare accessibility. This paper proposes a Network-based Covering Location Problem (Net-CLP) in the GIS environment, based on real world transportation networks with various travel thresholds.

Future work

Rigorous analysis of emerging data sources opens up a rich empirical context for social sciences research and

policy interventions. Many new exciting research developments continue to push the conceptual, theoretical, and technological boundaries that limit our ability to carry out this whole new field of action. Since most of new data and big data contain explicit or implicit geographical information, the emergence of such data signals a big opportunity for human geography. Economic geographers have been particularly interested in the spatial dimension of big data, which spans from the process of data production, data analysis and data management (Alasdair and Singleton 2015). For example, since the 1980s, economic geography has moved down to the local and up to the global. Relational approach and global network analysis have been widely applied to link the local and the global. Globalization and localization are identified as twin processes working together to reshape the global economic geography (Swyngedouw 2004). At the local level, micro level big data such as transactional data, customer data, firm level data and individual level data would provide richer and deeper understanding of local interactions for economic geographers. Moreover, big data can better depict the global networks of production, innovation, knowledge, capital and cities and the global–local interactions. In addition, big data can support the ambition of economic geographers to offer scientific supports for decision-makers.

Regions and cities are facing the daunting task of accommodating everyday human dynamics. Hence, smarter management strategies are needed in local operation and emergency response coupling with population growth trends. The massive data of human dynamics contains abundant knowledge about cities, regions and their citizens. Given the strain that larger populations will put on limited resources, the environment and infrastructure, understanding urban and regional big data can help optimize local operation, improve life quality and environment, and more efficiently deal with various emergencies. Robust, easy-to-use platform enabling effective exploration of new and big data is critical and will contribute to building capacity in seeking solutions for the social, economic, and environmental challenges facing our human settlements.

While a few urban regions are making progress in smart city operation, most are under-prepared and lack the resources and strategic planning tools needed to effectively address urban expansion. Domain

practitioners, researchers, and decision-makers need to conduct Real-Time Urban Surveillance to realize the tasks of smart city. However, they are facing great challenges due to the lack of computing infrastructures supporting big dynamic data collection and analytics in the real time. Understanding and analyzing the large-scale complex urban dynamic data is of great importance to enhance both human lives and urban environments. Policy practitioners, researchers, and decision-makers can all benefit from a new computing network and infrastructure enabling real-time on-site information abstraction from dynamic Big Data to conduct analytical tasks and make timely decision for urban emergency.

However, we also seek to caution against shifting towards a big-data-driven research too hastily and sometimes blindly. First, the interpretation of big data should be highly embedded in uneven and variegated local contexts (Graham and Shelton 2013). What big data can capture and reveal depends on the technology used, the context in which data are generated and the data ontology employed. Second, big data tend to be with multi spatial and temporal scales. Geographers have to structure big data with proper spatial–temporal scale. Third, big data are usually generated by private businesses and government through directed, automated or volunteered way (Kitchin 2013). Diversified provenance of big data increases the uncertainty due to its difficulties on quality control (Goodchild 2013). Limited access and uncertain quality brings about additional difficulties for structuring big data, which limits its utilization on scientific studies (Graham and Shelton 2013).

The multiple features of big data determines the advantages in visualizing network connection, recording individual behaviors and coupling the spatial–temporal and stock-flow information. Big data create many opportunities for urban and regional researchers. First, it can support the urban and regional development studies from a global–local perspective. Due to its power on visualization and simulation, big data can support the analysis of variegated flows like capital, commodities, information, knowledge, and labor. Second, behavioral analysis at the individual level may offer more details of firm heterogeneity in spatial strategy, helping understand the changing role of localized factors. Third, the fusion of spatial–temporal and stock-flow information may offer a way to monitor the supply and demand of energy, resources

and international trade. It would become the powerful data basis for environmental economic geography to couple the social-economic and eco-environmental systems (Yang et al. 2015). In this special issue, “CyberGIS and Spatial Data Science” argues that “interdisciplinary approaches combining rich and complex spatial data, analysis and models are highly demanded to ignite transformative geospatial innovation and discovery for enabling effective and timely solutions to challenging regional and urban problems”.

Recent advances in geocomputation techniques greatly enhance the abilities of social scientists to conduct large-scale data analysis. Such large-scale data analytics will stimulate the development of new computational models. In turn, these newly developed methods get adopted in real-world practice, forming a positive feedback loop. Enabling such urban and regional analysis over Cyberinfrastructure is particularly useful for projects such as a Smart City (Batty 2012), where this positive feedback loop will allow researchers and practitioners to test their models faster and scale them to a larger dataset, therefore producing more policy-relevant results (Wright and Wang 2011). To get deep insights from the new and big data, researchers must conduct iterative, evolving information foraging and sense making and guide the process using their domain knowledge. Iterative visual exploration is one key component in the processing, which should be supported by efficient data management and visualization tools. Therefore, urban and regional researchers demand a handful and effective visual analytics software which integrates scalable database and interactive visualization with powerful computational capability. A platform of real-time urban and regional system will enable more urban and regional scientists to develop specific technologies quickly. Domain users will be relieved from the burden of big data management and analytics, allowing them to focus on the design of research questions. The platform will advance a broad spectrum of real-world applications by enabling an extensive community of domain users to tackle the big data challenge.

However, traditional urban and regional studies are designed for small data, and we are therefore largely underprepared for the era of big data. This special issue thus seeks to figure out how urban and regional researchers can better cope with and extract useful information from the data deluge, and work towards a

more productive integration of big data with research paradigms. First, papers in this special issue emphasize that producing proper scales to structure the big data for empirical studies is fundamental. Although development of big data has highlighted the significance of coping with its messy nature, spatial-temporal scale issues remain to be solved. Second, the geography of big data is likely to be a research question for human geography and a proxy for technological development. Production and management of big data reflects the technological basis of certain region, and is closely related to localized factors. One statement that has arisen along with the emergence of big data is that with enough-volume data can speak for themselves, further signaling “the end of theory” (Kitchin 2013). This special issue points out that this naivety overlooks the fact that any analyses and interpretations of big data must be predicated on contextual or domain-specific knowledge (Ye et al. 2016). Third, we seek to make some theoretical development so as to translate the correlation derived from big data into the causation for theoretical hypothesis since data-driven approaches underestimate the role played by researchers in the analytical process. Finally, as we shift out emphasis towards big data-driven research, small data analysis should not be marginalized. Studies in this special issue show that small data studies can be tailored towards answering specific questions and thus complement big-data-based analysis (Shaw et al. 2016).

References

- Batty, M. (2012). Smart cities, big data. *Environment and Planning-Part B*, 39(2), 191–193.
- Chang, R. M., Kauffman, R. J., & Kwon, Y. (2014). Understanding the paradigm shift to computational social science in the presence of big data. *Decision Support Systems*, 63, 67–80.
- Gonzalez-Bailon, S. (2013). Big data and the fabric of human geography. *Dialogues in Human Geography*, 3(3), 292–296.
- Goodchild, M. F. (2013). The quality of big (geo) data. *Dialogues in Human Geography*, 3(3), 280–284.
- Graham, M., & Shelton, T. (2013). Geography and the future of big data, big data and the future of geography. *Dialogues in Human Geography*, 3(3), 255–261.
- Kitchin, R. (2013). Big data and human geography: opportunities, challenges and risks. *Dialogues in Human Geography*, 3(3), 262–267.

- Laney, D. (2001) 3D data management: Controlling data volume, velocity and variety. *METAGroup* <http://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf> (Accessed 29 Feb 2016)
- Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., et al. (2011). *Big data: The next frontier for innovation, competition, and productivity*. New York, NY: McKinsey Global Institute Report.
- Rae, A., & Singleton, A. (2015). Putting big data in its place: a regional studies and regional science perspective. *Regional studies, regional science*, 2(1), 1–5. doi:10.1080/21681376.2014.990678.
- Shaw, S., Tsou, M., & Ye, X. (2016). Human dynamics in the mobile and big data era. *International Journal of Geographical Information Science*, 30(9), 1687–1693.
- Swyngedouw, E. (2004). Globalisation or ‘glocalisation’? Networks, territories and rescaling. *Cambridge Review of International Affairs*, 17(1), 25–48.
- Wang, Z., Ye, X., & Tsou, M. (2016). Spatial, temporal, and content analysis of Twitter for wildfire hazards. *Natural Hazards*. doi:10.1007/s11069-016-2329-6.
- Wright, D. J., & Wang, S. (2011). The emergence of spatial cyberinfrastructure. *Proceedings of the National Academy of Sciences*, 108(14), 5488–5491.
- Yang, Z. S., Long, Y., & Douay, N. (2015). Opportunities and limitations of big data applications to human and economic geography: The state of the art. *Progress in Geography*, 34(4), 410–417.
- Ye, X. (2016). *Spatial econometrics*. New York: The Wiley-AAG International Encyclopedia of Geography.
- Ye, X., Huang, Q., & Li, W. (2016a). Integrating big social data, computing, and modeling for spatial social science, cartography and geographic information science (in press).
- Ye, X., Li, S., Yang, X., & Qin, C. (2016b). *Use of social media for detection and analysis of infectious disease in China*. China: ISPRS International Journal of Geo-Information (in press).