EDITORIAL　　　　　　　　　　　　　　　　　　　　　　**Open Access**

CrossMark

Big data and sustainable cities: applications of new and emerging forms of geospatial data in urban studies

Yeran Sun[1*] and Yunyan Du[2]

Introduction

With the development of Information and Communications Technology (ICT), new forms of geospatial data are emerging in recent years. Among the new and emerging forms of geospatial data, volunteered geographic information (or crowdsourced geographic information) and geosocial networking data are gaining much interest [1–4]. Sources of volunteered geographic information (VGI) include collaborative mapping projects (e.g., OpenStreetMap and WikiMapia) and social media (e.g., Flickr, Foursquare, Twitter, Facebook). As the term 'VGI' has been widely accepted and used, VGI refers to geographic information generated voluntarily. Collaborative mapping projects like OpenStreetMap (OSM) is intentionally contributed by people as data creators are always aware that they are creating geographic information; whereas geographic information, such as geotagged posts, images and videos, publicly shared in social media is likely to be unintentionally contributed by people as most of the data creators (social media users) might not be aware or never care that they are creating geographic information.

As the most up-to-date open access mapping platform in the world, OSM is playing an increasingly important role in both research and commercial applications. More and more researchers use OSM as a base map in their studies, and develop thematic maps (transport maps, disaster maps, etc.) or specific services (route planning). To offer users a complete and up-to-date understanding of local environment, a number of popular location-based services switch to OSM from commercial counterparts like Google Map, Yahoo Map and Bing Map. OSM is a crowdsourced map, which means the features and attributes are likely to be voted by the crowd. This enables OSM to maintain an acceptable level of data quality [5].

Unlike OSM data is intentionally contributed and voted by the crowed, social media data is likely to be unintentionally contributed and lack of votes from the crowed. As a result, it is hard to implement quality assurance of social media data, which leads to a high risk of bias. Apart from popular social media, other online social networks for certain communities (tourists, cyclists, hikers, researchers, etc.) and other forms of social networks (e.g., mobile phone networks) offer new forms of geospatial data. Those new forms of geospatial data all belong to geosocial networking data. Geosocial networking data tend to be 'big data' due to its large volume, unstructured forms and risk of bias. In spite of risk of bias, geosocial networking data gain an increasing attention because it has a high spatial accuracy and tracks human activities at the individual level. Therefore, VGI data and geosocial networking data will pave the way for a better understanding of some key issues or new issues in urban studies and environmental sciences. Compared to traditional geospatial data, geospatial big data require different data processing and analysis approaches.

Applications of geospatial big data in urban studies

In this section, we are briefly introducing the papers in the special issue. All these papers are using big data to address key social and environmental issues in the development of sustainable cities.

Urban resilience

Efficient disaster management and emergency planning make contributions to improving urban resilience. In disaster management and emergency planning, a fast response will reduce loss of life and property. New technologies facilitate a faster response to natural hazards and emergencies. One the one hand, new and emerging forms of data are integrated to establish warning and management systems for natural hazards. To share information in the shortest time possible, the Ministry of

* Correspondence: yeran.sun@glasgow.ac.uk
[1]Urban Big Data Centre, University of Glasgow, Glasgow G12 8RZ, UK
Full list of author information is available at the end of the article

Environment and Natural Resources of El Salvador [6] builds early warning systems for natural hazards monitoring in El Salvador. Using diverse spatial data (census data, agrometeorological data, telemetry data, satellite imagery data), they propose an integrated and comprehensive system and create a big database to manage and communicate information of the natural disasters situation. Mobile applications, such as weather hazard app and warning message app, are developed for the dissemination of information to people in real time.

Unlike some big natural hazards (floods, hurricanes, tropical storms and tsunamis) that are predictable, a variety of small emergencies are unpredictable. Accordingly, they are unlikely to be identified and communicated by monitoring and warning systems timely. In the era of citizen science, inhabitants can report real-time emergencies that they are seeing. Traditionally, inhabitants can make emergency phone calls to inform the occurrence of emergencies. However, as a vital piece of information the location is often not obtained automatically at all or with insufficient accuracy, thus limiting both the speed and the reliability of emergency response [7]. To address this issue, Šterk, et al. [7] develop a smart phone app for emergency to speed up and simplify the acquisition of the locational and other data. Apart from fast data acquisition, fast decision-making plays a vital role in fast response as well. Šterk, et al. [7] also develop a management system to help the dispatchers draw better and quicker decisions by estimating the time for emergency service units to reach the emergency site. In practice, the management system is used to find lost people, to establish new emergency response units where needed, and to manage resources in refugee camps.

Urban livability

Improving urban livability is a key issue in urban planning and development. As a fundamental part of development of a livable city, settlement is facing new challenges in accommodating a fast-growing urban population. Using diverse spatial data (census data, environmental data, survey data, geosocial networking data), Omar et al. [8] assess and map settlement suitability in Kirkuk city, Iraq. Specifically, they take account of a variety of environmental and socio-economic factors, and integrate and manage datasets with different forms and scales. Specifically, the environmental factors selected include distance from built up area, topographic (slope), distance from river, soil (bearing capacity), distance from road, and land use activities; while the socioeconomic factors selected include population density, land value, proximity to security area and same ethnic area. Additionally, the results of their study point out that the socioeconomic factors have an influence on the direction of the urban expansion in Kirkuk city.

In recent years, participatory urban planning is increasingly popular toward more livable cities. In traditional urban planning, there is a gap between desires of local inhabitants and goals of governments. To narrow the gap, participatory urban planning encourages inhabitants to participate in the planning and development of their neighborhoods and the entire cities. Advances in ICT enable involvement of inhabitants in urban management and development easily. Crowdsourced maps and smart phone apps make inhabitants have access to urban management and development. Typically, local people are able to assess the quality of the built environment in local areas. One the one hand, inhabitants can evaluate access to and service quality of amenities (public transport facilities, schools, healthcare facilities, retails, recreational facilities, restaurants, etc.) via online maps and social media. Sun et al. [9] exploit Yelp, a popular online social network for business owners (restaurants, stores, pubs, etc.), to evaluate service quality of restaurants, fast foods and bars. Yelp has a crowdsourced rating system, which facilitates the reviews of users on business owners. Sun et al. [9] map restaurants, fast foods and bars with ratings across a city, and afterwards identify clusters of high and low ratings in the city. On the other hand, inhabitants can express their real-time emotions that are associated with perceptions of the built environments or activities undertaken (work, rest, dine, etc.). Sayegh et al. [10] collect data for height, width, depth, loudness and light in a place (room or space) as well as individual subjective feeling via smart phone apps that make use of embedded sensors. By doing so, they measure unseen but detectable qualities of cities.

Sustainable transport

Sustainable transport is a key indicator for a sustainable city. Ridesharing (bike-sharing, car-sharing, hitchhiking, etc.) plays an important role in development of sustainable transport. Ridesharing impacts travel behavior of people [11, 12]. For instance, by increasing the use efficiency of cars, car-sharing makes contributions to reduction of energy consumption and greenhouse gas emissions. As a green and healthy transport, cycling is encouraged by governments all over the world. Apart from improving cycling facilities such as road conditions, separated bike lanes and traffic calming, launching bike-sharing programs contributes to an increasing usage of bicycle, and thus leads to a modal shift from motor vehicles to cycling. Additionally, due to health benefits of cycling, bike-sharing is likely to contribute to health promotion. As a result, popularity of ridesharing is reducing energy consumption, traffic congestions and carbon emissions in cities.

Compared to bike-sharing and car-sharing, hitchhiking is likely to be random and untraceable. Although hitchhiking

is older than bike-sharing and car-sharing, it is not studied as often as bike-sharing and car-sharing due to the lack of data. In recent years, a popular hitchhikers' website Hitchwiki allows hitchhikers to review hitchhiking locations with location ratings, waiting time, and comments. To get a better understanding of hitchhiking behavior, Vedernikov et al. [13] analyze rating and waiting time of hitchhiking locations in 10 countries. Moreover, they develop models to predict the quality and waiting time at hitchhiking locations according to locational characteristics related to road types.

Social and spatial inequalities

Moreover, reducing social and spatial inequalities is a main target of sustainable development. Socio-spatial polarization is associated with increasing influence of the central areas and decreasing influence of the peripheral areas. Places of high influence have a large number of inflows and outflows of goods, service, people or capital. Socioeconomic connections between places can be represented by movement of goods, service, people or capital. A large amount of flows of goods, service, people or capital between two places often indicates a strong socioeconomic connection between these two places. Places (neighborhoods, districts, cities, states, nations) and their socioeconomic interconnections can form a socio-spatial network. In a real-world socio-spatial network that is situated within a region such as a state, province or metropolis, cities of high influence are hubs with strong socioeconomic interconnections as they have the largest volume of inflows and outflows of goods, service, people or capital. At the same time, online social networks contain online social relationships ('friends' or 'followers') between people located in different places. Aggregating individuals' online social relationships to places (cities, provinces, states or countries) can generate an online socio-spatial network. In an online socio-spatial network, a large number of online social relationships between two places indicates a strong online social connection between the two places. Similarly, in a virtual socio-spatial network situated within a region, core cities are likely to be hubs with the strong interconnections as they have the largest population, economy and other resources. Although a city is likely to play a role with different magnitude of importance in a virtual socio-spatial network and a real-world socio-spatial network, its influence in the virtual network is likely to be positively associated with its influence in the real-world network.

Apart from flows of tangible resources (goods, service, people or capital), flows of intangible resources (information) provides a new perspective on socio-spatial inequalities and social cohesion at the national, regional or city levels. Online social networks offer a new approach to exploring spatial inequalities in urban and regional development. Extracting a virtual socio-spatial network from a popular social media in China, Ying et al. [14] analyze the intercity network situated in a region centered on Wuhan, named Wuhan "1 + 8" City Circle (WCC). Specifically, they propose and measure two indicators: degree of city cohesion and degree of city outreach, to quantify the relations within a city and the relations among different cities. The results indicate that in WCC the degree of cohesion and outreach in a city node were significantly positively correlated with the economic development level of the city.

The future of applications of geospatial big data in urban studies

In spite of challenges (data quality assurance, data privacy, etc.), geospatial big data will play an increasing role in urban studies. One the one hand, cities are considered complex systems and all the components are not working independently but interacting with each other. For the development of sustainable cities, urban planners and managers need to get a more comprehensive and up-to-date understanding of different aspects of cities. Diverse crowdsourced thematic maps will be developed to provide most up-to-date information of the local areas (streets or neighborhoods). This will facilitate fast updates on local accommodation, amenities and environmental quality in cities. On the other hand, as human-centric design has an increasing influence on the development of smart and sustainable cities, the efficient and effective acquisition of individual-level data is extremely vital. To harness power of citizen science for collecting individual data, an increasing number of location-based apps will be developed and used. In the near future, diverse location-based apps and embedded sensors will enable inhabitants to voluntarily collect real-time health-related information or built environment information align with their locations and activity types (e.g. work, rest, eat, sports).

In the future, new and emerging forms of geospatial data will be widely applied in addressing key issues in urban studies, including sustainability, livability, vulnerability, pollution, health, wellbeing, housing, transport and crime. Accordingly, both hardware and software need to be upgraded to cope with massive unstructured datasets, and standardization of geospatial data structure needs to pay more attention to new and emerging forms of geospatial data.

Authors' contributions

YS and YD design and develop the structure of the paper. YS writes the paper. Both authors read and approved the final manuscript.

Publisher's Note

Author details
[1]Urban Big Data Centre, University of Glasgow, Glasgow G12 8RZ, UK. [2]State Key Laboratory of Resources and Environmental Information System, Institute of Geographic Sciences & Natural Resources Research, Chinese Academy of Sciences, Beijing 100101, China.

References

1. Bakillah M, Liang S, Mobasheri A, Jokar AJ, Zipf A. Fine-resolution population mapping using OpenStreetMap points-of-interest. Int J Geograph Inform Sci. 2014;28(9):1940–63.
2. Mobasheri A, Sun Y, Loos L, Ali AL. Are Crowdsourced datasets suitable for specialized routing services? Case study of OpenStreetMap for routing of people with limited mobility. Sustainability. 2017;9(6):997.
3. Sun Y, Fan H, Bakillah M, Zipf A. Road-based travel recommendation using geo-tagged images. Comput Environ Urban Syst. 2015;53:110–22.
4. Sun Y, Mobasheri A. Utilizing Crowdsourced data for studies of cycling and air pollution exposure: a case study using Strava data. Int J environ res public health. 2017; 14(3): 274.
5. Senaratne H, Mobasheri A, Ali AL, Capineri C, Haklay MM. A review of volunteered geographic information quality assessment methods. Int J Geograph Inform Sci. 2016:1–29.
6. Rivera JY. Tools to operate and manage early warning systems for natural hazards monitoring in El Salvador. Open Geospat Data Softw Stand. 2016;1:9.
7. Šterk M, Praprotnik M. Improving emergency response logistics through advanced GIS. Open Geospat Data Softw Stand. 2017;2:1.
8. Omar NQ, Raheem AM. Determining the suitability trends for settlement based on multi criteria in Kirkuk, Iraq. Open Geospat Data Softw Stand. 2016;1:10.
9. Sun Y, Paule JDG. Spatial analysis of users-generated ratings of yelp venues. Open Geospat Data Softw Stand. 2017;2:5.
10. Sayegh A, Andreani S, Kapelonis C, Polozenko N, Stanojevic S. Experiencing the built environment: strategies to measure objective and subjective qualities of places. Open Geospat Data Softw Stand. 2016;1:11.
11. Sioui L, Morency C, Trépanier M. How Carsharing affects the travel behavior of households: a case study of Montréal, Canada. Int J Sustain Transp. 2013;7(1):52–69.
12. Shaheen S, Martin E, Cohen A. Public bikesharing and modal shift behavior: a comparative study of early bikesharing systems in North America. Int J Transp. 2013;1:35–53.
13. Vedernikov O, Kulik L, Ramamohanarao K. The Hitchhiker's guide to the pick-up locations. Open Geospat Data Softw Stand. 2016;1:12.
14. Ying S, Zhang N, Wang K, Zhao Z, Shen Z, Su D. Virtual network analysis of Wuhan 1+8 City Circle based on Sina microblog user relations. Open Geospat Data Softw Stand. 2017;2:2.