

ANALYSIS OF THE STATUS AND TREND OF GIS RESEARCH IN THE PAST TEN YEARS FROM 2010 TO 2019

Kai Hu^{1,*}, Zongming Dai¹, Jie Zheng^{2,3}, Huayi Wu^{2,3,*}

¹ Key Laboratory of Advanced Process Control for Light Industry, Ministry of Education, Jiangnan University, Wuxi 214122, China - (hukai_wlw@jiangnan.edu.cn, dzm_1995@163.com)

² The State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan 430079, China (zhengjie, wuhuayi@whu.edu.cn)

³ Collaborative Innovation Center of Geospatial Technology, Wuhan University, Wuhan 430079, China

Commission IV, WG IV/2

KEY WORDS: scientometric; local citation score; global citation score; network analysis; topic clusters; GIS research

ABSTRACT:

To understand the trends in a specific field is always an important issue for researchers in the field. In geographical information science (GIS) field, understanding the trends is also an open question for many researchers. In this paper, we conducted a scientometric analysis based on the selected GIS represented journals, provided by the work (Biljecki 2016). We use the data-driven method stemming from scientometric theories to help readers to obtain a direct impression of the field. Several interesting facts are found. The in-domain and out-domain institution and journals are compared. Structural analysis shows that classical research trend in GIS fields include Landsat, VGI, social sensing, change detection, spatial econometrics, and laser scanning.

1. INTRODUCTION

The research questions and application scenarios of Geographical Information Science (GIS) are continuously evolving. To understand the research status of the recent studies in the field is vital for GIS researchers. However, with the increase of publication amount in this GIS field, it will be overwhelming for researchers to quickly grasp the research focal problems and the preferred methodologies. How to properly deal with the problem is widely discussed.

Surveys, reviewing and outlining the publications by investigators apparently are the most common way to provide readers with the overview for research dynamics in the domain. However, this may cause inevitable drawbacks, because investigators may have their own backgrounds and preferences, thus making these surveys missing some potential important research directions. Publication amount explosion is not a problem in a single field, but also in many scientific fields. More importantly, these publications are becoming more and more available to more researchers.

In this situation, a discipline “scientometrics”, the so called “science of science”, has entered the moment of its spring (Fortunato *et al.* 2018). Many methodologies are kept bringing new ways to exploit the enormous publication datasets. Using the scientometric, the data-driven method may better help us to gain

new insight for the domain research status. Though, there may be some drawbacks when using the data-driven method by scientometric theories, they can be used as an alternative way that as the supplementary of the traditional survey.

In this paper, we conduct an analysis to analyze the representative academic journals of the GIS field, using numerical indicators and network structure. Firstly, we wish to use the numerical indicator numbers to provide the direct impression of the works in the past ten years. Then, the more profound knowledge by the structure analysis is explored.

The following sections are organized as follows: section 2 introduces the data and methodologies; section 3 describes the results of numerical numbers, section 4 discusses results by network structures; section 5 draws the conclusion.

2. DATA AND METHODS

2.1 Representative journal selection and article retrieval strategy

Due to different opinions on GIS, different experts may have different opinions. A previous work has summarized 20 representative GIS international journals selected by questionnaires (Biljecki 2016). By selecting the Science Citation Index-Expanded (SCI-E) and Social science citation index (SSCI) databases of Web of Science core collection, we collected

* Corresponding author

bibliographical datasets of GIS papers from 2010 to 2019, and based on the datasets, we conducted the analysis.

The search conditions are critical to the scientific nature of a bibliometric analysis, and the search conditions are as follows, so that the analytical data is scientific and reproducible.

(SO=("Annals of the Association of American Geographers" OR "Cartography and Geographic Information Science" OR "Computers & Geosciences" OR "Computers, Environment, and Urban Systems" OR "Environment and Planning B Planning Design" OR "Geographical Analysis" OR "GeoInformatica" OR "GIScience & Remote Sensing" OR "International Journal of Applied Earth Observation and Geoinformation" OR "International Journal of Digital Earth" OR "International Journal of Geographical Information Science" OR "ISPRS International Journal of Geo-Information" OR "ISPRS Journal of Photogrammetry and Remote Sensing" OR "Journal of Geographical Systems" OR "Journal of Spatial Information Science" OR "Journal of Spatial Science" OR "Photogrammetric Engineering and Remote Sensing" OR "Photogrammetrie, Fernerkundung, Geoinformation" OR "Photogrammetry, Remote Sensing and Geoinformation Processing" OR "Spatial Cognition and Computation" OR "Transactions in GIS"))

AND LANGUAGE: (English)

AND DOCUMENT TYPES: (Article)

Timespan: 2010-2019. Indexes: SCI-EXPANDED, SSCI.

Note that the current "Journal of Spatial Information Science" has not been indexed by SSCI or SCI by now, so there are only 19 GIS journals in the following analysis. The date of the search is November 1, 2019. The retrieval datasets might be fluctuated slightly because the database is continuously collecting the papers published. Through the above search criteria, we obtained 13,538 search results, including the author, title, keywords, abstracts, and cited references, institutions, countries, and journal.

2.2 Numerical indicators, network structures, and dual-map analysis

Many numerical indicators have been created to describe the activities of the academic field. Local Citation Score (LCS) and Global Citation Score (GCS) may be the direct way to figure out the difference between in-domain and out-domain importance of a certain paper. LCS and GCS are used by the famous bibliometric software HistCite (Garfield 2009). LCS stands for the citations the paper obtained from the current data collection, usually the collected bibliographical datasets using the search constraints which is far less than the whole WoS database. GCS stands for the citations the paper obtained from the whole WoS database. Thus, LCS stands the recognition from the current datasets, namely the current domain. GCS stands from the recognition from all-science community. Therefore, the in-domain and out-domain importance can be depicted. In the paper, Total Local Citation Score (TLCS) and Total Global Citation Score (TGCS) will also be used, standing for the sum for LCS and GCS of each paper in a certain collection.

LCS and GCS have high efficiency for researchers to understand the in-domain and out-domain importance. However, they do not

reveal the structural knowledge between the academic elements. When one paper cite two references, these two reference will have the co-cited reference relation. With the publication accumulates, the network will be built by these relations. The understanding for the structure may be more efficient to know only one cited paper.

3. ANALYSIS RESULTS –NUMERICAL INDICATORS

3.1 Yearly output

The amount of publication usually represents the level of activity in the field and is not necessarily positively related to the influence of domain research.

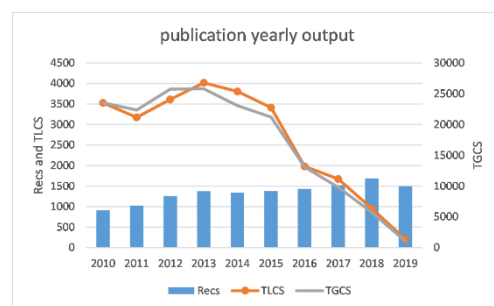


Figure 1. The number of papers published each year (where TLCS represents the values quoted within the dataset, representing the recognition within the domain, TGCS represents the value referenced by the entire Web of Science database, representing the recognition by all-science community)

From Figure 1, it can be seen that the TLCS has a local peak in 2013, which means that the work in 2013 has gained a great in-domain attention. It is worth noting that although the TLCS and TGCS are relatively low in the past three years, it is not that the researches in these years are not important. This is mainly because the published papers always take a certain amount of time to obtain citations. In another aspect, the TLCS and TGCS has similar pattern, also meaning that these work are similar important regarded by in-domain and out-domain readers.

To have a close look at the paper published in 2013, we can find several papers have gain great attention. "Spatial, temporal, and socioeconomic patterns in the use of Twitter and Flickr" has obtained the highest LCS of 47 (Li *et al.* 2013), "Integration of logistic regression, Markov chain and cellular automata models to simulate urban expansion" (Arsanjani *et al.* 2013) has the LCS of 43, ranked at the second place of the year 2013 high LCS paper list. "Virtual Geographic Environment: A Workspace for Computer-Aided Geographic Experiments" (Lin *et al.* 2013) received LCS of 38, ranked at the third place of the year 2013 high LCS paper list.

3.2 In-domain and out-domain importance of the journals

The 19 journals are ranked using three indicators, namely recs,

TLCS, and TGCS. Recs stands for the records, namely the amount of the publications. TLCS and TGCS are used for the paper collections, as shown in Table 1 and Table 2.

The journals *Computers & geosciences*, *ISPRS international journal of geo-information*, *International journal of applied earth observation and geoinformation* published the largest amount of articles from 2010 to 2019. But the journal *ISPRS journal of photogrammetry and remote sensing* and *International journal of geographical information science* have obtained the highest TLCS, 5611 and 3802, respectively. In term of TGCS, the journal *ISPRS journal of photogrammetry and remote sensing*, *International journal of applied earth observation and geoinformation* and *Computers & geosciences* have ranked at top three in the list.

From the list, we can tell that *ISPRS journal of photogrammetry and remote sensing* and *International journal of geographical information science* may be the most domain recognized journals.

Computers & geosciences, *ISPRS international journal of geo-information* published large amount of papers. Journal *Computers & geosciences* is becoming more and more recognized from outsider of the domain, because it ranks at the fourth place in the TLCS list, but third in the TGCS list.

4. ANALYSIS RESULTS -NETWORK STRUCTURE

The numerical indicators can help us to quickly identify the key elements in the academic activities. However, many important details are lost. Because any elements in the fields are not isolated, they are cooperating in certain forms. In this section, the network structure are mined and visualized with the scientometric software, CiteSpace. The structure include the sub-clusters in the co-cited references and the co-author network.

Journal	Recs	TLCS	TGCS
ISPRS journal of photogrammetry and remote sensing	1499	5611	32867
International journal of geographical information science	1131	3802	13971
International journal of applied earth observation and geoinformation	1604	3147	27341
Computers & geosciences	1990	2558	27180
Computers environment and urban systems	715	2023	9909
Transactions in GIS	571	1519	5001
Photogrammetric engineering and remote sensing	691	1374	8785
GIScience & remote sensing	418	1169	4253
International journal of digital earth	499	786	5604
Cartography and geographic information science	314	768	2728
ISPRS international journal of geo-information	1781	706	7045
Environment and planning b-planning & design	583	650	5645
Annals of the association of american geographers	480	612	11692
Geographical analysis	225	429	2171
Geoinformatica	255	365	2078
Journal of geographical systems	189	292	1750
Photogrammetrie fernerkundung geoinformation	182	253	1411
Journal of spatial science	268	218	1098
Spatial cognition and computation	143	76	799

Table 1. Journals ranked in domain metrics.

Journal	Recs	TLCS	TGCS
ISPRS journal of photogrammetry and remote sensing	1499	5611	32867
International journal of applied earth observation and Geoinformation	1604	3147	27341
Computers & geosciences	1990	2558	27180
International journal of geographical information science	1131	3802	13971
Annals of the association of American geographers	480	612	11692
Computers environment and urban systems	715	2023	9909
Photogrammetric engineering and remote sensing	691	1374	8785
ISPRS international journal of geo-information	1781	706	7045
Environment and planning b-planning & design	583	650	5645
International journal of digital earth	499	786	5604
Transactions in GIS	571	1519	5001
GIScience & remote sensing	418	1169	4253
Cartography and geographic information science	314	768	2728
Geographical analysis	225	429	2171
Geoinformatica	255	365	2078
Journal of geographical systems	189	292	1750
Photogrammetrie fernerkundung geoinformation	182	253	1411
Journal of spatial science	268	218	1098
Spatial cognition and computation	143	76	799

Table 2. Journals ranked in out-domain metrics.

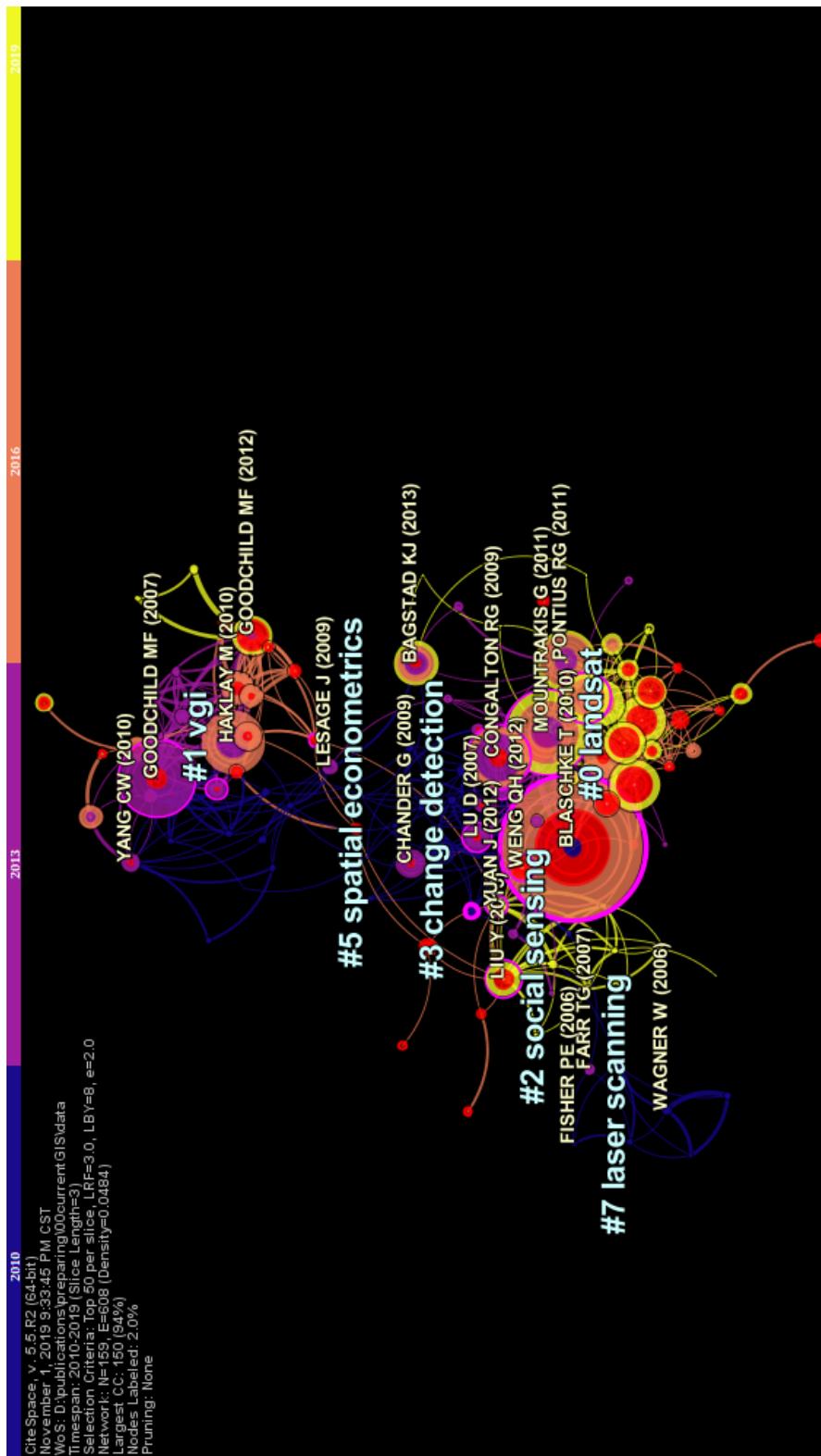


Figure 2. The co-cited reference network (default setting in the CiteSpace, every two year as a slice, “#” is the index of the cluster, the size of the nodes are proportional to the citation count of the reference. The red color in the node circle stands for the “burst pattern”. The purple color outside the nodes stands for the new branch of the research. Every cluster is set to show top three representative work, as there is limited visualization space)

Co-cited reference network is based on simple co-citation relation. When a paper cite two reference, these two references will have one co-citation relation. When the paper amount increase, the network based on the references will form. Based on the formed network, several community detection methods can be applied to detect the sub-clusters, thus forming highlighted #0, #1, #2, #3, #5, #7, as shown in Figure 2. The clusters are generated by using the betweenness centrality theory. With the nodes with highest betweenness removed, the clusters will be formed. Then based on the clusters, these clusters are labeled with terms generated by automatic method, namely Latent Semantic Index (LSI), Log-Likelihood Ratio (LLR), and Mutual Information (MI) (Jin-xia 2011). So these labeled words may not be seriously well defined terms, but they are computed as the most representative term for that cluster that can discriminate their work form other clusters.

Thus, we can see that Landsat, VGI, social sensing, change detection, spatial econometrics, and laser scanning are summarized. For cluster #0, Landsat is the representative word and cluster #0 is a huge cluster with many nodes inside. The representative work are the co-cited nodes “*Object based image analysis for remote sensing*” (Blaschke 2010), “*Support vector machines in remote sensing: A review*” (Mountrakis et al. 2011), and “*Death to Kappa: birth of quantity disagreement and allocation disagreement for accuracy assessment*” (Pontius and Millones 2011).

For cluster #1, VGI is the representative word, the full name of VGI is volunteered geographical information. The representative works include “*Citizens as sensors: the world of volunteered geography*”(Goodchild 2007), “*How good is volunteered geographical information? A comparative study of OpenStreetMap and Ordnance Survey datasets*”(Haklay 2010), and “*Citizens as sensors: the world of volunteered geography*”(Goodchild and Li 2012).

For cluster #2, “social sensing” is the representative word. The representative works include “*Social Sensing: A New Approach to Understanding Our Socioeconomic Environments*” (Liu et al. 2015), “*Discovering regions of different functions in a city using human mobility and POIs.*”(Yuan et al. 2012), and “*Remote sensing of impervious surfaces in the urban areas: Requirements, methods, and trends*”(Weng 2012). Note in this cluster, the forth represent work is a paper published in *Science*, titled “*Limits of predictability in human mobility*”(Song et al. 2010). It seems the paper in *Science* open the path for this cluster.

For cluster #3, “change detection” is the representative word, the representative works include the book “*Assessing the accuracy of remotely sensed data*” in 2009, “*A survey of image classification methods and techniques for improving classification performance*”(Lu and Weng 2007), and “*Summary of current radiometric calibration coefficients for Landsat MSS, TM, ETM+, and EO-1 ALI sensors*”(Chander et al. 2009).

For cluster #5, “spatial econometric” is the representative word, the representative works include “*A comparative assessment of decision-support tools for ecosystem services quantification and valuation*”(Bagstad et al. 2013), “*Introduction to spatial econometric*”(LeSage and Pace 2009),and “*Geospatial Cyberinfrastructure: Past, present and future*” (Yang et al. 2010).

For cluster #7, “laser scanning” is the representative word, the highlighted works are “The shuttle radar topography mission”

(Farr et al. 2007), “*Gaussian decomposition and calibration of a novel small-footprint full-waveform digitising airborne laser scanner*”(Wagner et al. 2006), and “*Causes and consequences of error in digital elevation models*” (Fisher and Tate 2006).

The cited work in this section may not belong the used 19 journals, but they can be regarded as the source of the key clusters, either contributing the key idea or methodologies. Totally 14 clusters are formed, but limited to the visualization space, only cluster #0,#1,#2,#3,#5,#7 are listed here, the rest clusters can be found the supplementary file.

5. CONCLUSION

In this paper, the classical GIS journals are collected from 2010 to 2019 and the scientometric analysis is performed using the numerical numbers, network structure analysis. Firstly, the local citation score (LCS) and global citation score (GCS) quickly outline the in-domain important and out-domain important journals,. Then the network-based methods co-cited reference networks are performed to obtain the understanding the active topics.

The work in 2013 has gained a great in-domain attention. Spatial, temporal, and socioeconomic patterns and virtual geographical environment are becoming important research topics. In term of journals, *ISPRS journal of photogrammetry and remote sensing* is the most representative journal and *International journal of geographical information science*. *Journal Computers & geosciences* seems to be more impactful in out domains than in GIS domain.

Cluster analysis shows that topics like VGI, change detections, social sensing, spatial econometrics, and laser scanning are the most kernel topics in the past few years. New topics like sensor networks, smart environments, 3D modeling, smart cities, IOT are also frequently discussed, but they are not well reflected in the clusters, it is because they need time getting cited in the network, thus being visible in the citation network.

ACKNOWLEDGEMENT.

This work is supported by Open Research Fund of State Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University (Grant no. 18104). The National Natural Science Foundation of China (no. 71904064) partially supports this research. The Natural Science Foundation of Jiangsu Province (no. BK20190580). The research is also supported by the 111 Project and Fundamental Research Funds for the Central Universities (Grant No: JUSRP11922).

REFERENCES

- Arsanjani JJ, Helbich M, Kainz W, et al.: Integration of logistic regression, Markov chain and cellular automata models to simulate urban expansion. *IJAEO* 21: 265-275, 2013.
- Bagstad KJ, Semmens DJ, Waage S, et al.: A comparative assessment of decision-support tools for ecosystem services quantification and valuation. *Ecosystem Services* 5: E27-E39, 2013.
- Biljecki F. A scientometric analysis of selected GIScience journals.

- International Journal of Geographical Information Science* 30: 1302-1335, 2016.
- Blaschke T. Object based image analysis for remote sensing. *Isprs Journal of Photogrammetry and Remote Sensing* 65: 2-16, 2010.
- Chander G, Markham BL, Helder DL. Summary of current radiometric calibration coefficients for Landsat MSS, TM, ETM+, and EO-1 ALI sensors. *Remote Sens. Environ.* 113: 893-903, 2009.
- Farr TG, Rosen PA, Caro E, et al.: The shuttle radar topography mission. *Rev. Geophys.* 45: 2007.
- Fisher PE, Tate NJ. Causes and consequences of error in digital elevation models. *Progress in Physical Geography* 30: 467-489, 2006.
- Fortunato S, Bergstrom CT, Borner K, et al.: Science of science. *Sci* 359: 2018.
- Garfield E. From the science of science to Scientometrics visualizing the history of science with HistCite software. *J. Informetr.* 3: 173-179, 2009.
- Goodchild MF. Citizens as sensors: the world of volunteered geography. *GeoJournal* 69: 211-221, 2007.
- Goodchild MF, Li L. Assuring the quality of volunteered geographic information. *Spatial statistics* 1: 110-120, 2012.
- Haklay M. How good is volunteered geographical information? A comparative study of OpenStreetMap and Ordnance Survey datasets. *Environment and Planning B-Planning & Design* 37: 682-703, 2010.
- Jin-xia Z. Documents Visibilization Analysis of Information Visibilization Based on the Citespace II [J]. *Information Science* 1: 022, 2011.
- LeSage J, Pace RK. Introduction to spatial econometrics. Editor (ed). Chapman and Hall/CRC. 2009
- Li LN, Goodchild MF, Xu B. Spatial, temporal, and socioeconomic patterns in the use of Twitter and Flickr. *Cartography and Geographic Information Science* 40: 61-77, 2013.
- Lin H, Chen M, Lu GN. Virtual Geographic Environment: A Workspace for Computer-Aided Geographic Experiments. *Annals of the Association of American Geographers* 103: 465-482, 2013.
- Liu Y, Liu X, Gao S, et al.: Social Sensing: A New Approach to Understanding Our Socioeconomic Environments. *Annals of the Association of American Geographers* 105: 512-530, 2015.
- Lu D, Weng Q. A survey of image classification methods and techniques for improving classification performance. *Int. J. Remote Sens.* 28: 823-870, 2007.
- Mountrakis G, Im J, Ogole C. Support vector machines in remote sensing: A review. *Isprs Journal of Photogrammetry and Remote Sensing* 66: 247-259, 2011.
- Pontius RG, Jr., Millones M. Death to Kappa: birth of quantity disagreement and allocation disagreement for accuracy assessment. *Int. J. Remote Sens.* 32: 4407-4429, 2011.
- Song C, Qu Z, Blumm N, et al.: Limits of predictability in human mobility. *Sci* 327: 1018-1021, 2010.
- Wagner W, Ullrich A, Ducic V, et al.: Gaussian decomposition and calibration of a novel small-footprint full-waveform digitising airborne laser scanner. *Isprs Journal of Photogrammetry and Remote Sensing* 60: 100-112, 2006.
- Weng Q. Remote sensing of impervious surfaces in the urban areas: Requirements, methods, and trends. *Remote Sens. Environ.* 117: 34-49, 2012.
- Yang CW, Raskin R, Goodchild M, et al.: Geospatial Cyberinfrastructure: Past, present and future. *Computers Environment and Urban Systems* 34: 264-277, 2010.
- Yuan J, Zheng Y, Xie X. Discovering regions of different functions in a city using human mobility and POIs Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining, 186-194, 2012.