Discussion on evolution of focal points of China’s urbanization: based on keywords analysis of papers from CNKI during 1992-2011

Jun Li¹,a, Suocheng Dong¹,b Zehong Li¹,c Qiliang Mao¹,d Yongbin Huang¹,e, Fei Wang¹,f, Peng Guo¹,g, Junni Wang¹,h
¹¹A, Datun Road, Chaoyang District, Beijing, 100101, China
a panghuanzhi@163.com, b dongsc3@163.com, c lizehong@igsnrr.ac.cn; d more1987@163.com, e huangyb.12b@igsnrr.ac.cn, f faye3611@126.com, g guop.12b@igsnrr.ac.cn, h wangjn.10b@igsnrr.ac.cn

Keywords: bibliometric; urbanization; research focus; keyword analysis

Abstract. Based on 24528 papers from China Knowledge Infrastructure during 1992-2011, we analyzed evolution of topics, theories, methods, and case regions. Results showed the hottest keywords changed from “township enterprises”, “peasant workers” to “urbanization agglomeration”, “land use” and “environment”. Evolution of hot topics is always tracing Chinese urbanization process. Theories are appearing dynamic, micro, and comprehensive. Dynamic econometric model has been main theory in industry structure research. Micro research on institutional modeling, urban agent’s environmental behaviors and peasant worker’s phycology has emerged. Modeling on land use and ecological environment effect has been more and more comprehensive. The hottest methods are mainly from statistical analysis. Evaluation on environmental effects of urbanization increases greatly in recent years while study on mechanisms of urbanization is still increasing slowly. Hottest case regions were mainly fast urbanization regions in East China including Jiangsu, Zhejiang and Guangdong, while West China was studied as a whole.

Introduction

Chinese urbanization rate rose from 26.41% in 1990 to 51.27% in 2011, at an average speed of 1.18% per year. During this fast urbanization process, Chinese urbanization research gained a great number of results. Systematic analysis on these fruits is helpful for researchers’ future studies.

Gu chaolin[1] grouped Chinese urbanization research fruits into 9 clusters: policies, features, dynamics, process, space, rural urbanization, comparison between Chinese and west countries, globalization and regional urbanization. Reviews of urban geography [2,3], economic geography[4] and human geography[5,6] also involved some summaries on Chinese urbanization research fruits. These reviews were usually abstract and used for grasping the sketches of Chinese urbanization research.

Based on some periodicals, some researchers made statistical analysis on Chinese urbanization research. Xu xueqiang[7] collected papers from main geographical periodicals, and summarized Chinese urbanization research in urban spatial structure, regional urban system, urbanization process, new scientific fields and methods.

Liu yungang[8] gathered 523 papers from Acta Geographica Sinica, Scientia Geographica Sinica and Geographical Research up to 2009 and summed up dynamic features of paper count, research type, methods, topics, and author affiliation. As the author himself pointed out that more systematic analysis was limited on database source and data process ability. Additionally, theory analysis was independent with topic analysis, so it was a bit hard for reader to see what theories were studied in given topic.

The China Knowledge Infrastructure (CNKI) was regarded as the most frequently used and most robust Chinese database for bibliometric studies[9]. Bibliometric analysis was used to analyze international research on urbanization based on SCI/SSCI database[10]. In this paper, based on 24528 papers from CNKI during 1992-2011, Excel VBA was employed to conduct bibliometric analysis on...
hot topics. For each main hot topic, we extracted main theories on that topic. Finally methods and case region analysis were made.

Data and Method

We built our bibliometric database using papers on urbanization from the CNKI. Time range was set from “1992” to “2011”. Topic item was filled with search term “urbanization”. This searching strategy allowed us to locate publications that contain “urbanization” in their titles, abstracts and/or keywords. Our bibliographic search resulted in 24,528 urbanization-related papers from the CNKI databases; we then extracted information about individual publications, including author name and affiliation, subject category, journal name, and publication year, keywords, cited frequency. Later, we performed a keyword analysis to reveal temporal evolution of topics, theories, methods and case regions in ecological environment.

Main Topics

Whole Keywords Analysis. Top 50 keywords were listed, of which had clear meaning were classified to 6 groups, according to the nearest rule. The first group was called regional difference group, including “urbanization process”. The second group named relationship between urbanization and economic development had “economic growth”, “economic development” and “rural economy”. The third group was about resource and environment, involving “land use”, “ecological environment”. The fourth group was the biggest family, with member of “industrialization”, “industry structure”, “Social security”, “local government”, “”, “township enterprises”, “Rural surplus labor”, “Rural urbanization”, “tertiary industry” and “Agriculture industrialization”. “Urbanization level”, “Central city” and “Urban agglomeration” constituted the fifth group. Suburbanization and gentrification group was the last, referring to “Peasant worker” and “land-lost peasants”.

Evolution of keywords indicated “Urban agglomeration”, “industry structure” and “ecological environment” were receiving more and more attention during 4 periods between 1992 and 2011. So it was useful to make detailed analysis about these keywords.

Evolution of Main Hot Topics. Firstly, we analyzed evolution of urban agglomeration (and related topics). Keywords containing urban agglomeration and related keywords were searched from keywords database, including “Urban agglomeration”, “urban system”, “Metropolitan Area”, “megalopolis cycle”, “megalopolis belts”, and “metropolitan interlocking region”.

Top 10 keywords were mainly basic concepts such as “Urban agglomeration”, “urban system”, “Metropolitan Area”, and “megalopolis cycle” except “Central China urban agglomeration” and “Changsha-Zhuzhou-Xiangtan urban agglomeration” which were the name of real urban agglomeration during 1992-2011. Besides these hottest keywords, “urban agglomeration planning”, “metropolis cycle economy”, “urban agglomeration community”, “urban agglomeration theory” and “metropolitan area planning organization” were main subfields in urban agglomeration field.

Evolution analysis showed that “urban system” ranked the first with a frequency of 20, nearly three times as much as the second, followed by “metropolitan area” and “metropolis cycle” during 1992-1996. Frequency of “urban agglomeration” transcended that of “urban system” and ranked the first, followed by “great metropolitan area” and “urban system” during 1997-2001. From 2002 to 2006, “urban agglomeration” got a frequency of 57 and ranked the top, leaving “metropolis cycle” at the second with a frequency of 25, followed further by “great metropolitan area” and “urban system”. During 2007-2011, frequency of “urban agglomeration” kept fast growing, leaving “metropolis cycle” and “great metropolitan area” behind, while frequency of “urban system” began to fell. Generally, “urban agglomeration” had much more frequency than “metropolis cycle” and “metropolitan area”, used most widely. Urban agglomeration and related concepts definition had caused hot discussion in Chinese urban geographical community[11]. Actually, this problem involved urban population definition, administrative zoning and statistical indexes. Not only researchers but also governmental officers should take part in discussion and make clear definition of these concepts in time.
Besides the basic concept keywords, more and more keywords from real urban agglomeration name came out, as well as higher and higher rank. During 1992-2001 this group of keywords only made up 9.38% of total frequency. Starting from 2002, represented with “Central China urban agglomeration” and “Changsha-Zhuzhou-Xiangtan urban agglomeration”, frequency of these keywords gained fast increase and took up 14.29% of total frequency. During 2007-2011, besides frequency of “Central China urban agglomeration” and “Changsha-Zhuzhou-Xiangtan urban agglomeration” grew to 14 from 6 and 5, respectively, more keywords sprang up in this group and the frequency proportion reached 24.07%. Since great difference between urban agglomeration, most urban agglomeration came from East and Middle China, only a few such as “North Bay urban agglomeration”, “urban agglomeration of upward reaches of Yellow River”, “Chengdu-Chongqing urban agglomeration” from West China. Considering its great importance on Chinese environment, urban agglomeration in West China needed more attention.

Besides more study on real urban agglomeration, research subfields were enriched. 1992-2001 was concept establishment stage with few keywords like “urban system planning”. Whole features of urban agglomeration were studied during 2002-2006, including “urban agglomeration planning”, “urban agglomeration economy”, “urban agglomeration theory”, “urban agglomeration feature”, “urban agglomeration pivot” and “urban agglomeration community”. Studies on inner structure of urban agglomeration in forms of “urban economic space”, “differentiated urban agglomeration”, “governmental relationship in urban agglomeration community”, “multiple center metropolitan area” and “metropolitan area governance” began to spring during 2007-2011. At the same time, ecological environment research on urban agglomeration emerged, including “urban agglomeration heat island”, “urban agglomeration community” and “urban agglomeration tourism”. As more data accumulation, research subfields on urban agglomeration would be richer and richer.

Secondly, we analyzed evolution of urbanization mechanism (and related topics).

Analysis on keywords containing “industry” indicated that “industrialization” emerged 833 times, leaving others far behind. The rest could be divide into two groups-countryside urbanization and new industrialization. The first group included “countryside industrialization”, “town industry” and “countryside industry”. The second group contained “new urbanization”, “industry park” and “new industrialization way”.

Keyword evolution showed that 4 of top 11 keywords were related to “countryside urbanization”, taking up 55.43% of total frequency. National industrialization, represented with keywords like “industrialization”, “industrialization and urbanization”, “industrialization level” and “industrialization process”, made up 60.17% of total frequency during 1997-2001. Featured with new industrialization, “industry park”, “new industrialization way”, “industry park construction” and “special industry park” were new hot topics, with an proportion of 23.78% during 2002-2006. New industrialization study evolved into micro level, featured with “industry heritages”, “industrial land” and “industry agglomeration”.

Actually, evolution of keywords related to industry accompany with Chinese industrialization. In the early 90th last century, township enterprises flourished, resulting in countryside industrialization. Though this golden time ended in the late 90th, yet development of township enterprise had accumulated necessary marketing education, capitals and human resources. So national industrialization could take up this baton and went to push Chinese industrialization[12]. After 2000, press from environment and resources had been heavier and heavier, Chinese government decided to implement new industrialization. Responding to central government’s advocate, industry parks sprang up in one night within nationwide. However, industry park construction caused amazing land waste. So research hotspot turned to “industry land” from “industry park”. Following this trend, ecological environmental effect would be the next focus as global climate changes.

Thirdly, we analyzed evolution of industry structure and related topics.

Keyword sort for all keywords manifested that “industry structure” and “tertiary industry” experiencing descending. However, analysis on “industry” showed difference between them. During 1992-2011, frequency of “industry structure” increased, yet its frequency seemed grew slower that
other hotter keywords and its rank fell. On the contrary, frequency of “tertiary industry” was experiencing lasting decreasing after 2006. Other keywords like “industry cluster”, “industry agglomeration” and “dominant industry” were within top 10 hottest topics.

Though “tertiary industry” was the hottest keyword during 1992-2001, frequency of “agriculture industrialization” and “industry structure” grew faster than “tertiary industry”. Specially, from 1997 to 2001, keywords about “agriculture industrialization” took up 31.23% of total frequency of top 10 keywords, sitting in the first place actually. During 2002-2006, “agriculture industrialization” and “industry structure” transcended “tertiary industry”, were the top 2 hottest keywords, especially frequency of “agriculture industrialization” had nearly doubled, as high as 104. Besides “industry structure”, related keywords, including “industry cluster” and “industry agglomeration”, had been hotter and hotter, and especially “industry transfer” seemed to be the next focus as Chinese government emphasized more on industry structure adaptation.

As for detail industry, “real estate industry” received the most attention. During 1997-2001, researchers concerned a lot about information industry. From 2002 to 2006, “high tech industry” became the focus. Culture-turn emerged during between 2007 and 2011, represented with hottest keywords like “culture industry”, “circulation”, “sports industry”, and “creative industry”. With the come of knowledge era, knowledge production, circulation and transference have been hotspots in international research and would be future Chinese research focuses.

Thirdly, we analyzed evolution of industry structure and related topics.

Generally, keywords related to land and water resource took up 0.15% and 0.06% of total keyword frequency respectively, constituted the core topics of ecological environment research during Chinese urbanization process. “Heat Island Effect” and “Climate Change” made up 0.09% and 0.07%, respectively, belonged to the latest hot spots. Further analysis on keywords related to land resource (KLR) indicated that land use was the core topic of land resource research, with a proportion of 23.74% of total KLR frequency. Eighty two keywords with clear meaning of ninety keywords related to land use (KLU) were divided to three groups—nonstandard, pattern, process and mechanism. Proportions of frequency of keywords in these three groups were 19.44%, 27.88%, 52.69%, respectively. Though results above showed that process and mechanism research had been the mainstream of land use research. Yet after checking keywords involved in process and mechanism research, it was found that most keywords were basic indicators such as transfer matrix, dynamic index, and dynamic degree and so on, rarely related to evolution process and modeling research. Detailed study was also made to keywords in pattern research, and found land use degree, structure, type, and benefit were the hottest topics in pattern research of land use. Only a few keywords, such as “Land Use Niche”, “Intensive Land Use”, “Land Use Distribution Pattern” and “Land Use Evaluation System” were related to comprehensive research on ecological effects of land use during Chinese urbanization process.

Top 10 keywords in four periods, 1992-1996, 1997-2001, 2002-2006, and 2007-2011, were obtained from keywords database. Land use was always the hottest topic in all four periods, especially during 2002-2006, six of the top 11 hottest keywords relating to land use topic, indicating researchers’ great concerns on land use problems during Chinese fast urbanization process. “Water Resources” had high frequency and ranked 5th during 1997-2006, a little descending after that. Since water resource was one of key factors in analysis of dynamic mechanism of Chinese urbanization, it would receive lasting attention in near future. “Heat Island Effect” had a lasting increase of frequency and reached the second position in the hottest keywords order with a frequency of 86 during 2007-2011, 2.58 times as many as that during 2002-2006. Further analysis made on “Heat Island Effect” showed heat island effect factors had been the newly hottest topic, together with former hot topics including heat island intensity, heat island warming rate, and heat island spatial distribution. As Chinese megalopolis were growing fast[14] more and more attention would be paid to heat island intensity, heat island warming rate, heat island spatial distribution and heat island effect factors on a larger urban scale.
Frequency of “Climate Change” had grown from 10 during 2002-2006 to 75 during 2007-2011, an increase of 6.5 times, one of the most concerned keywords in recent years. Relative humidity, average temperature, annual precipitation, and annual sunshine hours were main topics in climate change research. Limited to data access, citizen health resulted from climate change received unexpected low attention, compared to its significance. As development of health geography[15], this topic would become hotter and hotter. Similarly, “Landscape Architecture” had a frequency of 63 during 2007-2011, while rarely seen during 2002-2006. This kind of sudden growth was not accidental, and Yang[16] asserted that human civilization ecological turning, climate and environment problems, globalization and localization, fast urbanization, growing spiritual demand from lust for material were all backgrounds of this kind of explosive increase. However, whether it could represent one new research direction was till to be proved by time. Besides these hottest keywords, many others were in list, such as “Municipal Solid Waste”, “Urban Rain Flood”, “Air Pollution”, “Heavy Metal Pollution”, “Unground Water Pollution”, “Urban Non-Point Source Pollution” and “Rural Surface Pollution”. On one hand, as pointed out above, maybe because of data loss, research on these topics tended to ignore environment effects on citizens’ health. On the other hand, since urban region was an ecological system, how to take these factors as a whole for analysis needed more exploration in future.

Main Theories

Based on keyword database, we searched theories corresponding specified hot topic using both this topic and theory words, including “theory”, “model”, “law”, “method” and “analysis”.

Mechanism Theory. During 1992-1996, keywords about theory of urbanization mechanism were some basic concepts. From 1997 to 2001, “Chenery model”, “Todaro Migration Model” and “Stepwise Regression Model” emerged. During 2002-2006, theories became rich in three subfields. The first subfield, industry agglomeration, depended on “industry agglomeration theory”, “new trade theory”, “New classical economics” and “counting model”. The second subfield, industrialization, was mainly on system model besides “General Equilibrium Model”, including “Principal Component Analysis”, “Grey Relational Grade Analysis” and various predictive models. The third subfield, government, focused on “public finance theory” and “Game Theory”. Two features were noticeable during 2007-2011. The first was econometric models, famous with “VAR model” and “Panel Data model”. The second were some attempts in institution modelling, representative with “dual differentiation” in analyzing effect of Household Registration System on surplus labor transference and “Hierarchical Linear Model” in research on institutional interaction between urbanization and service industry. Totally, evolution of mechanism experienced from concept and theory introduction to empirical models application. Dynamic mechanism research is becoming mainstream and institutional modelling needs still further exploring.

Empirical Theory. The period of 1992-1996 was featured with forecasting models on urban population, centered on “Multiple Linear Regression” and “Differential Equation”. During 1997-2001, “Chenery model” and “Regression model” were main urbanization level research models and “rank-size rule” was applied in urban system study. From 2002 to 2006, theories and models on urbanization level enriched, including “Logistic model”, “error correction model”, “Northam curve”, “Lewis Model” and “Grey forecasting model”, with a few concepts on urban agglomeration. Between 2007 and 2011, researchers made more reflection on Chinese fast urbanization, in order to evaluate Chinese urbanization quality subjectively, comprehensive models on urbanization level measurement became popular, including “entropy AHP” and “Multiplication Model”. Additionally, more and more urban agglomeration theories emerged, most from ecological environment fields, including “collaborative symbiotic model” and “urban agglomeration heat island model”. Totally, Logistic model had gained lasting modification in urbanization measurement study, as well as applied in more and more fields such as land use and ecological environment effect during urbanization. Centered on land use, comprehensive models involving resource, economic and environment would be one direction on urbanization measurement.
**Suburbanization Theory.** The real study on peasant worker was starting from 1997, and “marginal man” came out and some interviews were launched on their “modernity”. During 2002-2006, transformation of urban village and suburbanization caused land expropriation and land-loss peasant problems. AHP was typical method for study on these topics while peasant worker research was mainly some policy analysis. From 2007 to 2011, focus of peasant worker research turned to peasant worker’s urbanization, coming with theories like “Dual Economic Model” and “ERG demand model”. Moreover, topics itemed from social problems, such as “Human Resource Development of Rural Migrants” and “Cultural Psychology”, became new hot topics. It was expected more theories would be introduced from sociology as more and more researcher paid attention to social problems with peasant worker.

**Main Methods**

Based on urbanization keyword database, keywords containing “method”, “analysis”, “model”, “technique”, and “metrics” and so on, were found and divided as 1992-1996, 1997-2001, 2002-2006 and 2007-2011 four periods(61, 103, 257, and 561 of frequency, respectively). Considering source and application of method, these methods were classified into four groups, including statistics, system, mechanism, and evaluation. Generally, these four groups had total frequency of 755, statistics accounting for 65.17%, system making for 13.11%, mechanism taking up 11.79% and evaluation occupying 9.93%, respectively. So it was evident statistical method was the widest method used in Chinese urbanization papers. Further analysis on keywords related to statistical method indicated that statistical methods could be divided into metric and spatial analysis methods. In detail, metric methods clustered around multivariate statistical analysis, including Principal Component Analysis, Factor Analysis and Correlation Analysis. Modified methods, such as Error Correction Model (ECM), Spatial Econometric Model, and Panel Data Analysis had been used in more and more papers. These methods took consideration of resource factors, leaving environmental factors ignored. So how to introduce environmental factors into models was one challenge for researchers in future. Spatial Analysis methods were mainly about GIS analysis, reflecting the truth that GIS had been widely applied in urbanization pattern analysis[17]. Yet it was also noticed that “Gradient Analysis” and “Buffer Analysis” were the most frequent keywords relating to GIS analysis, which showed application of GIS in urbanization papers was still basic, lack of comprehensive analysis on ecological environment problems. It was predicted that as citizens became more and more concerned about their living environment quality and health equality, it would be hotter and hotter to make analysis of urban environment health and urban landscape pattern using GIS. Within system method group, Analytical Hierarchy Process (AHP) and Grey System Analysis were the widest used methods. As for mechanism method group, Game Analysis had great potential in analyzing different social agents’ behaviors on environment problem caused by renewal of inner city and expansion of outer city. Most of evaluation methods were from ecology in urbanization papers, including famous Energy Analysis[18] and Ecological Niche[19]. Since urbanization research was showing a turning to ecological environment, progress in ecology analysis methods would be doomed to attract more and more attention.

Statistical and mechanism methods dominated in 1992-2001(Fig.1). During 2002-2006, methods were greatly enriched, attributed to fast development of system and evaluation methods, including AHP and Grey System Analysis from system methods, and R/S analysis from evaluation. On the contrary, there was only a little increase in number of traditional statistical and mechanism methods. Represented by STIRPAT Model, ECM and Co-integration Analysis, development of evaluation and statistical methods had a great progress in 2007-2011. Considering topic analysis above, it could be inferred that more and more applied resource-environment-economic comprehensive models would be developed for evaluation and simulation of ecological environment effects.
Main Case Regions

Generally, the hottest provincial name keywords (PNKs) were sorted in an east-middle-west order. Yet as for East China, “East China” lagged behind many eastern PNK. Specially, “Jiangsu”, “Guangdong” and “Zhejiang” were always the top 3 hottest PNK. Considering these three provinces were among the most mature urbanization regions in China[20], it could be found that researchers tended to select mature urbanization regions as their case regions. At the same time, it was clear to see areal differentiation of urbanization in East China. Local urbanization research was the mainstream of urbanization study in East China. Some typical urbanization regions also received researchers’ great attention besides provincial case study. Perl River Delta was at the forefront of Reform & Opening, first acceleration in urbanization, attracting researchers’ interests. After ranking 3rd during 1992-1996, position of “Perl River Delta” began to drop, along with changed research spots. Yangtze River Delta was opened a step later than Perl River Delta, so its rapid urbanization process fell behind a little and “Yangtze River Delta” ranked 5th at its peak position during 2002-2006, a little later than “Perl River Delta”. Beijing-Tianjin-Hebei Region urbanization was less developed than both Perl River Delta and Yangtze River Delta[21]. “Beijing-Tianjin-Hebei” ranked far behind “Perl River Delta” and “Yangtze River Delta”. Similar situation happened to “Middle China”, “Yangtze River Region” and “Longhai Railway”. Contrary to “East China”, “West China” was ahead of many western PNKs, which indicated researchers tended to take West China as a whole on urbanization. Objectively, areal differentiation of urbanization in West China might be lower than East China, since these regions all together sat in a lower urbanization stage. Yet, concerning great natural differences between different western provinces, different western provinces had different urbanization processes. So it is still necessary to make more special research on urbanization of local regions in West China[22].

Further, it was noticed that “West China” reached its peak 5th during 1997-2001, which was the right time of proposal of Western Development. Similarly, “Northeast” rose during 2002-2006 when Northeast Revitalization was heated discussed. Also, “Middle China” ranked 9th on the ground of Middle Rises during 2002-2006. All these demonstrated national policy had an important role in researchers’ case region selection.
Conclusions

After analyzing evolution of topics, theories, methods and case regions based on keyword database from CNKI, main conclusions were as follows.

Evolution of hot topics is always tracing Chinese urbanization process. Following this law, future hot topics include urban agglomeration (governmental relationship in urban agglomeration community and ecological environment effect of urban agglomeration unit), industry structure (industry cluster, industry agglomeration and industry transference) and ecological environment (urban environmental health and urban ecological environment effect).

Theories are appearing dynamic, micro, and comprehensive. Dynamic econometric model has been main theory in industry structure research. Micro research on institutional modeling, urban agent’s environmental behaviors and peasant worker’s psychology has emerged. Modeling on land use and ecological environment effect has been more and more comprehensive.

Statistical analysis is still main method. Methods based on strict mechanism analysis increase slowly. As more and more attention is paid to ecological environment, it is expected methods from ecology and environment would be richer and richer.

Fast urbanization regions are main case regions, including special regions in national development planning during different periods. Considering West China’s great importance in environment effect, West China urbanization, especially that of fragile environment region, needs more attention.

Acknowledgements

This study is supported by the funding from National Natural Science Foundation of China (NSFC: D011203; D010201).

References


