Human Dynamics in Smart Cities

Series Editors: Shih-Lung Shaw · Daniel Sui

Xinyue Ye
Hui Lin *Editors*

Spatial Synthesis

Computational Social Science and Humanities

Springer

Human Dynamics in Smart Cities

Series Editors

Shih-Lung Shaw, Department of Geography, University of Tennessee, Knoxville, TN, USA

Daniel Sui, Department of Geography, Ohio State University, Columbus, OH, USA

This series covers advances in information and communication technology (ICT), mobile technology, and location-aware technology and ways in which they have fundamentally changed how social, political, economic and transportation systems work in today's globally connected world. These changes have raised many exciting research questions related to human dynamics at both disaggregate and aggregate levels that have attracted attentions of researchers from a wide range of disciplines. This book series aims to capture this emerging dynamic interdisciplinary field of research as a one-stop depository of our cumulative knowledge on this topic that will have profound implications for future human life in general and urban life in particular. Covering topics from theoretical perspectives, space-time analytics, modeling human dynamics, urban analytics, social media and big data, travel dynamics, to privacy issues, development of smart cities, and problems and prospects of human dynamics research. This will include contributions from the participants of the past and future Symposium on Human Dynamics Research held at the American Association of Geographers annual meeting as well as other researchers with research interests related to human dynamics via open submissions. The series invites contributions of theoretical, technical, or application aspects of human dynamics research from a global and interdisciplinary audience.

More information about this series at http://www.springer.com/series/15897

Xinyue Ye · Hui Lin
Editors

Spatial Synthesis

Computational Social Science and Humanities

Springer

Editors
Xinyue Ye
Department of Landscape
Architecture and Urban Planning
Texas A&M University
College Station, TX, USA

Hui Lin
School of Geography and Environment
Jiangxi Normal University
Nanchang, China

ISSN 2523-7780 ISSN 2523-7799 (electronic)
Human Dynamics in Smart Cities
ISBN 978-3-030-52733-4 ISBN 978-3-030-52734-1 (eBook)
https://doi.org/10.1007/978-3-030-52734-1

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Introduction: Spatial Synthesis in Computational Social Science and Humanities

1. Towards Computational Spatial Social Science and Humanities

As Goodchild et al. (2000) illustrate, “changes in the space and place of peoples and nations have profoundly affected the spatial organization of the social, the economic, the political, and the cultural—the key domains of focus of the social sciences.” Space and place are central across social science and humanities disciplines to serve as both inputs and outputs of empirical and theoretical investigations (Dezzani 2010). Goodchild (2020) also states “in essence the particular form of integration that is so central to GIS practice is what we might term spatial integration.” Different from traditional social science and humanities, computational social science and humanities adopt computation as the vital enabling methodological foundation and platform (Cioffi-Revilla 2014). The Internet and cellular data networks significantly change our mode of communication and reshape the formation of networked groups which were previously strongly constrained by distance and location, releasing the power of social interactions and group assembly across a much larger territory. Furthermore, Lazer et al. (2009) announce the coming age of computational social science, because we are entering the life in the network digitally captured to form comprehensive pictures of both individuals and communities.

The past decade has witnessed the dramatic growth of computational social science and humanities research spurred by the increasingly available fine-scale and human-centered spatial (spatiotemporal) data. The computing environment for geo-visualization, geo-simulation, geo-collaboration, and human participation has also been developed to assist various computer-aided research tasks (Lin et al. 2013). The following major trends have been identified:

1. Spatial social science and humanities research has shifted from a data-scarce to a near real time data-rich environment. The availability of unprecedented data sources over space, time, and social networking would facilitate the modeling of

individuals' behavior and the outcomes of such model across spatiotemporal scales, deeply rooted in both geographic landscape and social network. For instance, Ye et al. (2019) integrate the spatial method and social network analytics to model the scope and sources of online transactions and quantify the driving forces, based on online transactions at the city level. A powerful analytical framework for identifying space-time research gaps and frontiers is fundamental to the comparative study of spatiotemporal phenomena upon configuration of various intertwined relationships. For example, novel research questions can be generated when we can systematically query the dynamic virtual and physical dimensions across multiple scales in socioeconomic modeling, transportation analysis, and disaster response (Ye and Rey 2013; Li et al. 2017; Wang and Ye 2017).

2. As the space-time data accumulate, the rich details of spatiotemporal dynamics in computational modeling remain largely unexplored because of many binding constraints for scientific advancement such as the challenge of intensities of data computing and very large geo-referenced dynamic databases (Shaw ad Ye 2019). In addition, such 24/7 unstructured social data needs special methods (Batty 2020). The revolution of computing and information technology has further blurred the boundary and definition of disciplines and applications. The increasing affordability of computing cost and lowered learning curve have also accelerated big spatiotemporal analytics studies at a growing rate.
3. The integration has been gradually realized among and across conceptualizations, analytical methods, and open-source software environments across disciplines of social science and humanities. Such an integration is needed to respond to the new data and computing environment (Liu et al. 2019). Human dynamics has been emphasized from the geospatial dimension within the context of mobile and big data era (Shaw et al. 2016). A virtual geographic environment has also been proposed as a computer-aided workspace for geographic experiments and analyses involving both the physical and human dimensions (Lin et al. 2013). By integrating environmental psychology theory and geospatial artificial intelligence, a framework of virtual geographic cognition experiment has been further developed to model and simulate human activity and urban context data (Zhang et al. 2018).

2. Synthesis and Convergence

Themes of Social Science and Humanities are increasingly relevant to convergence and synthesis across multiple disciplines as well as the data, computing, interactive, and collaborative environments. Annual International Symposium of Spatially Integrated Social Science and Humanities have been held ten times to promote such practice. This book is born from the most exciting and dominant themes in Spatial Synthesis: Computational Social Science and Humanities research in China. As a first English book of such kind, it spans most social science and humanities disciplines as well as computational science. This book is a comprehensive text on spatial and computational social science and humanities research. The development of more powerful computing technology, emerging big and open data sources, and theoretical perspectives on Spatial Synthesis has revolutionized the way in which we investigate social science and humanities. Given the pace of change and prominence of human-centered computing and spatial social science/humanities research, a summary of the principles and applications of such research is urgently required and will be of great value. In the foreword of this book, Batty (2020) notes that "the continued miniaturization of computers to the point where we are now using them personally in real time to organize our lives has led to many new ways of sensing and delivering data about our social behaviours", while Goodchild (2020) highlights core reasons supporting convergence and synthesis: "the pressing challenges faced by the earth cannot be solved by one single discipline, and hence need the collaborative work between computing experts and domains scientists for broader perspectives".

This book contains research and contributions from scholars across China and the World. The main principles and applications of spatial social science and humanities over the past decade in China are reviewed. The book provides fundamental information that will help to shape future research. This book will allow researchers, students, and policy-makers worldwide to learn about the significant achievements and applications of spatial social science and humanities research within China.

3. Spatial Synthesis in Humanities, Regional Science, and Urban Science

This volume is the Human Dynamics in Smart Cities book series and is composed of 25 chapters. After the forewords by Academicians Michael Batty and Michael Goodchild, the following chapters cover a variety of interesting and timely topics on Spatial Synthesis for Computational Social Science and Humanities. The chapters focus on three aspects: humanities, regional science, and urban science according to their different roles, pertinent issues, and corresponding solutions, as below:

Spatial Synthesis in Humanities: According to Hu et al. (2020), the official history, chorography, and family trees form the memory of China as a nation. They construct a multilevel architecture of Family Tree Geographical Information System (FTGIS) by incorporating modern geospatial information technologies into the research on family trees. Lu and Zhang (2020) build Historical Geographic Information System to promote the research of Chinese history including literature, maps, remote-sensing images, and archeological relics. Digital Historical Yellow River system is also developed to contain: (1) high-precision three-dimensional micro-geomorphology; (2) fusion scheme of historical hydraulic engineering and terrain model; (3) restoration of the three-dimensional shape of river channel; (4) simulation and demonstration of motion process in historical period of surface water; (5) reconstruction of rainfall characteristics in historical periods; and (6) river-water management methods in historical periods (Pan et al. 2020). Based on CHGIS (China Historical Geographic Information System) and CBDB (China Biographical Database) and mapping tools, Xu Y (2020) visualizes the trajectory, activities, and social networks of Tang Xianzu, a Chinese playwright of the Ming Dynasty. Measuring the cultural effects on demographic behaviors and outcomes is difficult because such influences are challenging to quantify, Xu H (2020) integrates biomarker data and small area estimation techniques to identify the spatial variation of cultural tolerance. How to effectively and efficiently protect the traditional Cave-Dwelling village is crucial for cultural heritage conservation. Dang et al. (2020) adopt Cultural Landscape Gene theory to analyze landscape features of cave-dwelling village in Wudinghe River Basin and examine its cultural values. Through the perspective of the cultural space, Shi (2020) explores the popularity of the Taiwanese ballad music form characterized by the mixed-race influences from Japan, by integrating the geographic information system and qualitative interviews.

Spatial Synthesis in Regional Science: Gu et al. (2020) systematically review the recent advance on Spatial Demography from the angles of differentiation and isolation, birth and death, migration and urbanization, regional population forecast, population and the environment, as well as analytical methods and application. Yang and Li (2020) document the previous studies for the air and high-speed railway networks at different spatial and temporal scales, based on the various configuration of complex network in the weighted network. Quite a few USA-based residential property owners have gotten financial support because of energy-efficient products and services through the programs such as the Property Assessed Clean Energy, Pan (2020) computes the economic influences of the residential energy efficient programs based on a metropolitan input–output model. Zhang et al. (2020) estimate the influences of changes in industrial structure, energy total factor efficiency, and energy structure on changes in carbon emission (CO_2) at the provincial level in China, using exploratory spatial data analysis and spatial panel econometric models. Gui et al. (2020) conduct a visual analysis of smart cities and big data management in the Yangtze River Delta region, based on company registration information for 30 years. Using both ordinary least square and geographically weighted regression, Gao and Chen (2020) analyze the driving forces of land urbanization in China at the county level in 2000 and 2015, finding land urbanization experienced an average increase by 2.77% annually during this period with an obvious north–south disparity. Population growth, economic development, industrial structure, city/county features, and geographical location are found to be significant factors shaping the geographical disparities of land urbanization. Qin et al. (2020) use the Chinese General Social Survey data to explore the geographical patterns and driving forces of intergenerational education mobility via intergenerational mobility indices and geographically weighted regression model. Adopting a representative sample of volunteered geographic information crawled from Sina

Weibo and Baby Back Home, Yang and Sui (2020) analyze the spatial distribution of child beggars and missing children in China, respectively.

Spatial Synthesis in Urban Science: To illustrate how geospatial data might be influenced by the rapid advance of artificial intelligence, Zhao et al. (2020) examine three geospatial spoofing cases: the game player trajectories generated by bot, the tweeted fake locational information, and simulated image of place made by a deep learning algorithm. Jiang (2020) promotes a complex network angle on the wholeness to better understand the nature of order or beauty for sustainable design, which helps to reduce the mystery of wholeness and enables us to appreciate Alexander's wholeness philosophy in fine and deep structure. Zhou and Peng (2020) develop an analytical framework of behavior research in China, fundamental to comparative study as well as dynamic and predictive research. Gao et al. (2020) adopt travelers' perception towards city space through bloggers, tweets, pictures, and videos, examining tourists' perceived images of city center, historical community, and traditional water town of Shanghai. Shen (2020) investigates the patterns of distance-based accessibilities for various housing types associated with surrounding community facilities for four counties in North Carolina, U.S.A. In addition, taking the transportation network companies vehicle GPS trajectories in Shenzhen as a case, Tu et al. (2020) design a data-driven framework to uncover on-demand shared mobility pattern. Zheng et al. (2020) review the applications of eye movement experiments in humanities, social science, and geospatial cognition. Furthermore, two experiments are conducted based on goal searching strategy and indoor wayfinding.

4. Conclusion

Academia, decision-makers, and citizens have progressively realized the necessity of modeling, simulating, and analyzing social phenomena based on large-scale computing, in order to transform our understanding of our lives, organizations, and societies at this point in the human history (Lazer et al. 2009). Noting that chapters in this book do not cover the full scale of computational spatial social science and humanities research, the following research avenues are also noteworthy. First, there is a need to develop a systematic and theoretical framework to characterize such methodological integration in reflecting the multifaceted nature of human dynamics and social complexity. Second, data-challenged depressed communities deserve the emerging strand of study because ignoring the coexistence of data-rich and data-poor environments would lead to possibly biased model results that are meaningless and harmful in policy implementation. To address fairness in big data analytics of social science and humanities offers us new opportunities in understanding the world and harnessing data science for social good. While we may not be able to mention all relevant studies in this short introductory piece, this edited volume is among the efforts to promote spatial synthesis in human and social dynamics studies, towards a new generation of research environments and tools to contribute to a deeper understanding of the geographic world.

Xinyue Ye
Hui Lin

Acknowledgement The work has been supported by School of Geography and Environment, Jiangxi Normal University.

References

Batty, M. (2020). Foreword I: Charting computational social science from a spatial perspective. (This volume).

Cioffi-Revilla, C. (2014). *Introduction to computational social science*: Principles and Applications, Springer: New York.

Dang, A., Zhao, D., Chen, Y., & Wang, C. (2020). *Conservation of cave-dwelling village using cultural landscape gene theory*. (This volume).

Dezzani, R. (2010). Spatially integrated social science. In B. Warf (Ed.), *Encyclopedia of geography*. doi: 10.4135/9781412939591.n1070

Gao, J. & Chen, J. (2020). *Demystifying the inequality in urbanization in China through the lens of land use*. (This volume).

Gao, J., Ma, J., Li, J., & Wang, L. (2020). *Studies on tourists' city space images*. (This volume).

Goodchild, M. F. (2020). Foreword II: Convergence and synthesis. (This volume).

Goodchild, M. F., Anselin, L., Appelbaum, R. P., & Harthorn, B. H. (2000). Toward spatially integrated social science. *International Regional Science Review*, 23(2), 139–159.

Gu, H., Lao, X., & Shen, T. (2020). *Research progress on Spatial Demography*. (This volume).

Gui, R., Chen, T., & Wu, Z. (2020). *Spatial visualization and analysis of the development of high-paid enterprises in the yangtze river delta*. (This volume).

Gui, Z., Wang, Y., Li, F., Tian, S., Peng, D., & Cui, Z. (2020). High performance spatiotemporal visual analytics *technologies and its applications in big socioeconomic data analysis*. (This volume).

Hu, D., Cheng, X., Lu, G., Wen, Y. & Chen, M. (2020). *The China family tree geographic information system*. (This volume).

Jiang, B. (2020). *A complex-network perspective on alexander's wholeness*. (This volume).

Lazer, D., Pentland, A. S., Adamic, L., Aral, S., Barabasi, A. L., Brewer, D., & Jebara, T. (2009). *Life in the network: the coming age of computational social science*. Science: New York, NY, 323(5915), 721.

Li, M., Ye, X., Zhang, S., Tang, X., & Shen, Z. (2017). *A framework of comparative Urban trajectory analysis. Environment and planning B*. doi: 10.1177/2399808317710023

Lin, H., Chen, M., & Lu, G. (2013). Virtual geographic environment: a workspace for computer-aided geographic experiments. *Annals of the Association of American Geographers*, *103*(3), 465–482.

Liu, X., Xu, Y., & Ye, X. (2019). Outlook and next steps: Integrating social network and spatial analyses for urban research in the new data environment. In *Cities as spatial and social networks* (pp. 227–238). Springer, Cham.

Lu, Y & Zhang, P. (2020). *GIS for Chinese history research*. (This volume).
Pan, Q. (2020). *Economic impact analysis for an energy efficient home improvement program*. (This volume).
Pan, W., Su, R., Man, Z., Zhang, L., He, M., & Han, L. (2020). *Digital historical yellow river*. (This volume).
Qin, K., Luo, P., Lu, B., & Lin, Z. (2020). *Analysing spatial patterns of intergenerational education mobility in China*. (This volume).
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.
Shen, G. (2020). *Accessibility of residential houses to commnuity facilities*. (This volume).
Shi, C. (2020). *Digitalized enka-style Taipei*. (This volume).
Tu, W., Wei, C., Zhao, T., Li, Q., Zhong, C., & Li, Q. (2020). Uncovering online sharing vehicle *mobility patterns from massive GPS trajectories*. (This volume).
Wang, Z. & Ye, X. (2017). Social media analytics for natural disaster management. *International Journal of Geographical Information Science*. doi: 10.1080/13658816.2017.1367003
Xu, H. (2020). *Quantifying spatial variation in aggregate cultural tolerance*. (This volume).
Xu, Y. (2020). *Visualizing classic Chinese literature*. (This volume).
Yang, H. & Li, Y. (2020). *Complex network theory on high-speed transportation systems*. (This volume).
Yang, X. & Sui, D. (2020). Can social media rescue child beggars? (This volume).
Ye, X. & Rey, S. J. (2013). A framework for exploratory space-time analysis of economic data. *Annals of Regional Science*, *50*(1): 315–339.
Ye, X., Lian, Z., She, B., & Kudva, S. (2019). Spatial and big data analytics of E-market transaction in China. *GeoJournal*, 1–13.
Zhang, F., Hu, M., & Lin, H. (2018). Virtual geographic cognition experiment in big data era. *Acta Geodaetica et Cartographica Sinica*, *47*(8), 1043.
Zhang, J., Li, J., & Wang, X. (2020). *Exploring the dynamics of carbon emissions in China via spatial-temporal analysis*. (This volume).
Zhao, B., Zhang, S., Xu, C., & Liu, X. (2020). *Spoofing in geography: Can we trust artificial intelligence to manage geospatial data?* (This volume).
Zheng, S., Chen, Y. & Wang, C. (2020). *Application of eye-tracking technology in humanities, social sciences and geospatial cognition*. (This volume).
Zhou, S. & Peng, Y. (2020). *Spatial-temporal behavior analysis in Urban China*. (This volume).

Contents

Part I
Foreword

Chapter 1
Foreword I: Charting Computational Social Science from a Spatial Perspective

Michael Batty

Although digital computers emerged in the first half of the 20th century, the idea of computation had been deeply embedded in science and philosophy from the Enlightenment, certainly from the Renaissance on, and indeed as far back as the Greeks. When computers were invented however, it took on a new meaning in that everything which computers were able to do first depended upon reducing a problem to its digital fundamentals and then combining and recombining its elements using logics, arithmetic, and algebras. This is the contemporary notion of 'computation' in contrast to the term 'computer' which is reserved for the hardware on which such computation takes place. In fact computation has come to dominate the myriad of applications that that define the scope that computers can address, and slowly but surely over the last 80 years, the term 'computational' has been appended to many areas as computers increasingly penetrated social and economic life, well beyond their original applications in science. In the late 20th century, the term began to be applied to various of the social sciences. For example, 25 years ago, it was used by Hummon and Farajo (1995) in their paper on computational sociology. In the late 1980s, David Mark and his colleagues at the National Center for Geographic Information and Analysis used the term in many conversations and in 1994, The Centre for Computational Geography was set up by Stan Openshaw at the University of Leeds (http://www.ccg.leeds.ac.uk/). This led directly to the term Geocomputation which is still widely used to this day and whose history I recalled in an editorial to mark the 21st anniversary of the first conference (Batty, 2017).

But back to social science. The publication of the path-breaking book by Epstein and Axtell (1996 *Growing Artificial Societies: Social Science from the Bottom Up*) was a wonderful demonstration of how computation could be employed to simulate many features of contemporary communities and their histories, showing all

M. Batty (✉)
CASA, University College London, London, UK
e-mail: m.batty@ucl.ac.uk

X. Ye and H. Lin (eds.), *Spatial Synthesis*, Human Dynamics in Smart Cities,
https://doi.org/10.1007/978-3-030-52734-1_1

the features of complexity science—segregation, agglomeration, income distribution, spatial clustering, reflecting new methods and ideas ranging from emergence, fractals, positive feedback, power laws, historical accident, path dependence and so on. It introduced agent-based modelling in contrast to much of the aggregative modelling in the social sciences that had preceded it. Thus computational social science came to define the use of computers to enable simulations to be extended to quite large systems but more specifically to methods that enabled many different kinds of logics other than classical algebras to be applied to diverse problems in social and economic domains. A decade after the millennium, the field had matured to the point where representation as well as simulation set the boundaries on its scope, reflected in Claudio Cioffi-Revilla's (2010) review of the field where he defines "the main computational social science (CSS) areas are automated information extraction systems, social network analysis, social geographic information systems (GIS), complexity modelling, and social simulation models." What marks this definition is that CSS is methodologically self-conscious in that although it deals with the social and economic domains extending as far as the behavioural sciences of individual economic and social decision-making, it does not presume to extend our knowledge of these substantive systems. CSS is not focussed on developing new social science theory per se although it may be based on demonstrating how we can develop new methods for validating theory, and in this process, there is often some focus on articulating ways of measurement and simulation which may eventually lead to new and better theory.

For the last 10 years, there has however been a sea change with respect to CSS. The continued miniaturisation of computers to the point where we are now using them personally in real time to organise our lives has led to many new ways of sensing and delivering data about our social behaviours. This is data that is captured and often delivered, often analysed and acted upon in near real time. It is data that is 'big' in the sense that individual behaviours are being captured 24/7 and it is voluminous in size. It often requires special and very different multivariate techniques and methods to even represent, store and access it as this is data that is largely unstructured. Unlike official Census data, it is not made to measure and often requires the powerful tools coming from what is now called data science for exploring whether significant patterns exist within it. In this sense, the focus changes within computational social science to developing much more inductive methods, methods that seek to extract patterns which ultimately build up to new hypotheses, rather than develop simulations which seek to test these hypotheses. Of course, the scientific methods in CSS are no different from any other positive philosophies which always depend on a fusion of inductive and deductive perspectives.

In this book, Hui Lin and Xinyue Ye have put together an interesting collection of papers that deal with a very wide range of computational approaches to not only the social sciences but also the humanities. What distinguishes this set of papers other than its extent is the fact that the expertise of the editors in geospatial analysis is brought to bear on the various papers. Computational geography as we alluded to above is an additional theme that runs throughout the collection and this serves to ground the various chapters in quite well-developed GIS technologies. In fact as

Cioffi-Revilla (2014) notes, social GIS (geographic information systems/science) is key to his more catholic definition of CSS and this is certainly the stance taken by the editors.

The book is divided into three parts, all dealing with communicating a synthetic knowledge of computation in the humanities, regional science, and urban science in that order. The first part deals with the humanities covering the structure of geographic information using ideas about hierarchy, the use of GIS in Chinese historical research, the visualisation of Chinese literature, cultural landscapes, conservation, and archaeological perspectives. The second and third parts deal with changes in scale, to some extent from national concerns to the regional and then the urban. Spatial demography, network theory as in transportation, economic impact analyses, carbon emissions, the locations of firms, analysis of social and economic structure through visualisation, inequalities, educational mobility and poverty are all key dimensions in the papers developed in this section. There is a stronger quantitative dimension to the papers here where spatiotemporal modelling and visualisation are widely developed.

The book then changes tack to deal with cities at the urban scale. Illusions and twists in geographic analysis introduce this focus and then the tenor changes to complex networks, spatiotemporal behaviour, imageability in cities, community facilities, mobility, and tracking. All of these papers are written using spatial tools which emphasise visualisation of complex data sets. In fact most of the data introduced in what are a set of strongly empirical papers do not really fall into the class of big data per se. But the tools of simulation and visualisation in computational social science are well developed here and potential readers will be able to gain a real sense of how geospatial analysis can be used in CSS to great advantage. Many of the examples relate to different spatial scales in Chinese cities and regions and this provides a fascinating explanation of how wide such science is and how it is being developed for important advances in our understanding of explanation and prediction in social systems from a spatial perspective.

References

Batty, M. (2017). Geocomputation. *Environment and Planning B: Urban Analytics and City Science, 44,* 595–597.

Cioffi-Revilla, C. (2010). Computational social science. *WILEY Interdisciplinary Reviews: Computational Statistics, 2*(3), 259–271.

Cioffi-Revilla, C. (2014). *Introduction to computational social science: Principles and applications.* New York: Springer.

Epstein, J. M., & Axtell, R. L. (1996). *Growing artificial societies: Social science from the bottom up.* Cambridge, MA: The MIT Press.

Hummon, N. P., & Fararo, T. J. (1995). The emergence of computational sociology. *Journal of Mathematical Sociology, 20*(2–3), 79–87.

Chapter 2
Foreword II: Convergence and Synthesis

Michael F. Goodchild

In their book *Convergence of Knowledge, Technology, and Society* Roco et al. (2013; see also NRC 2014) argued that the history of science has been one of swings between divergence and convergence. In the divergence phase specialization flourishes, with limited interaction between specialties, while in the convergence phase the barriers between the specialties begin to weaken, and science advances through the sharing of expertise and interest between specialties. The US National Science Foundation has recognized the importance of convergence in today's scientific enterprise, defining it as "integrating knowledge, methods, and expertise from different disciplines and forming novel frameworks to catalyze scientific discovery and innovation" (https://www.nsf.gov/od/oia/convergence/index.jsp).

There are several reasons for believing in the importance of convergence and synthesis at this point in the history of science. First, the problems faced by humanity are arguably more challenging than they have ever been, as the planet becomes more crowded and its ability to sustain life is under increasing threat. Second, science today is by nature collaborative, requiring specialists in statistics, computing, and other cross-cutting disciplines in addition to the expertise in particular domain sciences that is required by the problem at hand; the days when a lone investigator working in a single discipline could isolate and study a problem and derive significant knowledge from it are probably gone. Third, many of the practices of academia are centripetal, drawing a scientist into a real or imagined core of his or her discipline; it follows that positive effort is required to encourage and reward broader perspectives.

Yet anyone who has followed the history of geographic information systems (GIS) since their inception in the mid 1960s will be familiar with an earlier version of the convergence argument. In building his school of landscape architecture at the University of Pennsylvania in the 1950s and 1960s, Ian McHarg argued that expertise in a

M. F. Goodchild (✉)
University of California, Santa Barbara, CA, USA
e-mail: good@geog.ucsb.edu

X. Ye and H. Lin (eds.), *Spatial Synthesis*, Human Dynamics in Smart Cities,
https://doi.org/10.1007/978-3-030-52734-1_2

number of disciplines—in ecology, hydrology, climatology, geology, soil science—was essential to good landscape architecture. Thus in staffing the school he made sure to hire experts in each of these areas, and to insist that each of them be dedicated to synthesis, to making the school more than the sum of its disciplinary parts (McHarg 1969, 1996). In developing a plan, each discipline's contribution could be visualized as a layer of knowledge, one of the stack of layers that now graces the front cover of many GIS textbooks and the left margin of many GIS software products. In short, GIS has always claimed a role in integration and in supporting interactions across the boundaries of disciplines.

In essence the particular form of integration that is so central to GIS practice is what we might term spatial integration. Driving a spike through all of the layers will intersect the same location on each layer, and have the effect of integrating each discipline's data for the point in space that is represented by the spike. The ecologist's information about that point can now be coupled with information from the geologist, the hydrologist, the climatologist, and the soil scientist. This argument raises space to a central role in integration or convergence. In the late 1990s I proposed that the National Science Foundation fund a Center for Spatially Integrated Social Science, an investment in the infrastructure of the social sciences that would explore and demonstrate the value of location in enabling conversations and synthesis between the social sciences. The center was established at UCSB in 1999, and for five years it organized a series of programs: workshops, software development, learning resources, examples of best practice, improving access to tools, and search for data based on geographic location (csiss.org; Goodchild and Janelle 2004; Goodchild et al. 2000). The work of the center continues today in UCSB's Center for Spatial Studies (spatial.ucsb.edu).

It is easy to see how this argument for the role of geographic space in integration can be extended to time, and to any discipline that deals with phenomena distributed in space and time. A compelling argument can even be made that space and time are unique in this respect; that the processes studied largely independently in the domain sciences need be integrated only when it is necessary to study their joint impacts on a location at a specific time.

Early progress on the development of GIS was slow, due at least in part to the heavy demands that it placed on very limited computing resources. Combining vector data required that intersections be computed between the lines and areas depicted on each layer, and it was not until the late 1970s that algorithms, methods of indexing, and computing resources had advanced to the point where this was feasible and reliable (Esri's PIOS and Harvard's ODYSSEY both emerged at about this time). Instead, a common work-around was to represent each layer in raster, using a common, co-registered raster for each layer, despite the old adage that "raster is faster but vector is correcter". Several systems emerged in the early 1970s to implement what was in practice a very simple raster overlay operation. Today vector overlay algorithms are fast and reliable, and many of the early raster-overlay systems disappeared or were absorbed by the industry leaders.

But another twist to this argument has emerged in recent years. Raster systems were always seen as single-scale, working at a fixed spatial resolution, yet today we have access to a great variety of raster data at a wide range of resolutions, and interesting advances have been made recently in multi-scale analysis, combining layers at different resolutions. The technology of discrete global grid systems (DGGS; Sahr et al. 2003) allows the planet's surface to be divided into tiles that are approximately equal in size and shape, at a hierarchy of levels of resolution, with each level nesting within the level above. DGGS are superbly elegant ways of integrating multi-scale data.

This new book on spatial synthesis is one more proof of the value of this approach. Each chapter takes one area of the social sciences and humanities and shows how a spatial approach can result in significant advances in knowledge. It should be of great value to anyone interested in pursuing this approach, or in developing new tools to support it, or in developing courses that can empower students. Congratulations to the organizers and editors; I look forward very much to seeing it in print.

References

Goodchild, M. F., & Janelle, D. G. (2004). *Spatially integrated social science.* Oxford, New York.

Goodchild, M. F., Anselin, L., Appelbaum, R. P., & Harthorn, B. H. (2000). Toward spatially integrated social science. *International Regional Science Review, 23*(2), 139–159.

McHarg, I. (1969). *Design with nature.* Garden City, NY: Natural History Press.

McHarg, I. (1006). *A quest for life.* Wiley, New York.

NRC (National Research Council). (2014). *Convergence: Facilitating Transdisciplinary Integration of Life Sciences, Physical Sciences, Engineering, and Beyond*. Washington, DC: The National Academies Press.

Roco, M. C., Bainbridge, W. S., Tonn, B., & Whitesides, G. (Eds.). (2013). *Convergence of knowledge, technology and society: beyond convergence of nano-bio-info-cognitive technologies.* New York: Springer.

Sahr, K., White, D., & Kimerling, A. J. (2003). Geodesic discrete global grid systems. *Cartography and Geographic Information Science, 30*(2), 121–134.

Part II
Spatial Synthesis in Humanities

Chapter 3
The China Family Tree Geographic Information System

Di Hu, Xinghua Cheng, Guonian Lü, Yongning Wen, and Min Chen

3.1 Family Tree and GIS

3.1.1 Family Tree

A family tree (also called genealogy) is important historical material. The official history, chorography and family trees constitute China's national history. A family tree systematically documents a clan with the same ancestor. A large amount of historical information about individuals, families, clans, society, ethnology, customs, economy, peoples, geography, population and culture is contained in a family tree (Ge 1996; Wang 2006).

Family trees have great value that is mainly reflected in four aspects: cultural relics, literature, education and rooting (Ge 1996; Wang 2006). First, a family tree is a cultural relic; some family trees have existed for more than 1,000 years. Some family trees have been edited or commented on by celebrities.

Second, the family tree is an important form of literature that can provide ample and important data for many research fields, including studies of the family, history, humankind and surnames. The family tree is a kind of useful material for researchers

D. Hu · X. Cheng · G. Lü (✉) · Y. Wen · M. Chen
School of Geography Science, Nanjing Normal University, Nanjing 210023, China
e-mail: gnlu@njnu.edu.cn

D. Hu · G. Lü · Y. Wen · M. Chen
Key Laboratory of Virtual Geographic Environment, Ministry of Education, Nanjing Normal University, Nanjing 210023, China

Jiangsu Center for Collaborative Innovation in Geographical Information Resource Development and Application, Nanjing 210023, China

X. Cheng
Department of Land Surveying and Geo-Informatics, The Hong Kong Polytechnic University, Kowloon, Hong Kong, China

X. Ye and H. Lin (eds.), *Spatial Synthesis*, Human Dynamics in Smart Cities,
https://doi.org/10.1007/978-3-030-52734-1_3

who wish to investigate feudal thought and the family system of China. Furthermore, national historical events, the lives of celebrities, and histories of minorities and families are recorded in family trees with different levels of detail. These records are important reference materials for historical research.

Third, family trees have educational value. Generally, family tree records include parental instructions, the regulations of a clan and the laws of a family. These records reflect traditional Chinese virtues. Investigating a family tree enables a researcher to trace a family's heritage and development and reproduce a glorious history that can greatly inspire the descendants of the family.

Fourth, family trees have rooting value. An increasing number of overseas Chinese people are interested in identifying their ancestors. In this sense, family trees can be regarded as an important reference that provides important evidence of family history. The origin and lineage of a family are important content that makes a family tree essential material for such genealogical research efforts.

A great advantage can be found in Chinese family trees. First, Chinese family trees have a long history, originating from the pre-Qin period and continuing to the present. Family trees have existed over several centuries, and some have existed for nearly 1,000 years. Second, numerous family tree records exist in China, providing researchers with important source material. Some family trees have been well preserved, which enables researchers to extract useful information from them. Third, with increasing awareness of tracing roots and ancestors, family trees are continually consulted by both the public and research institutes. Therefore, it is important to investigate and mine the information contained in Chinese family trees.

3.1.2 GIS and Family Tree Research

Geographical information system (GIS) is a computer system that stores, manages, analyzes, expresses and displays geographic information about geolocation-related phenomena (Goodchild 2009). In the several decades since its initial development, GIS has been applied in all walks of life, including environmental protection (Goodchild 1993), hydrologic modeling (Devantier and Feldman 1993), land use analysis, agriculture, public health (Nykiforuk and Flaman 2011; Higgs 2004), transportation and urban planning (Harris and Elmes 1993). Current hot topics in GIS research include three-dimensional GIS, service-oriented GIS, digital globe (Goodchild 2018) and smart cities (Roche 2014; Degbelo et al. 2016) and so on.

GIS has enabled a focus on the organization, management and spatial analysis of geographic information about natural phenomena over recent decades. However, few studies have put effort into studying geographic information about the humanities and social sciences. In recent years, applying GIS to solve problems related to the humanities and social sciences has become increasingly popular, and GIS has been widely used in fields such as history, linguistics, criminology and economics. The concept of spatially integrated humanities and social sciences has been proposed (Harris 2009; Rumsey 2009; Goodchild and Janelle 2010). Many research institutions related to GIS and the humanities and social sciences have been established, and corresponding conferences have been held successfully. Moreover, a lot of databases and information systems have been established, such as Chinese Civilization in Time and Space (CCTS) (Liao and Fan 2012; Academia Sinica 2002), the China Historical Geographic Information System (CHGIS) (Harvard CGA 2001), the Historical GIS Database of Cotton Textile Industry on the Songjiang Region in Late Ming China (Billy 2011) and the Spatial History Project at Stanford University (White 2010).

The family tree can be considered a new source of geographic information because of its typical spatial-temporal characteristics. Family trees contain extensive information about individual, family and clan activities that occur in the context of specific spatial-temporal scenarios. The spatial information includes the birthplace and death place of an individual, the location of a grave and the site of an event. The temporal information includes the times of births, deaths, migrations and other events.

GIS has a variety of functions, including spatial-temporal information modeling, analysis, expression and display. These functions are highly useful for family tree research and use. Family tree information can be stored, managed, analyzed, expressed and displayed from spatial and temporal perspectives using GIS.

3.1.3 Concept and Objectives of Family Tree GIS

The concept of the Family Tree Geographical Information System (FTGIS) was proposed by Lü et al. (2009). The FTGIS emphasizes the importance of obtaining and mining family tree information about the spatial-temporal distribution and migration of clans and then expressing the spatial-temporal clan pedigree visually. The FTGIS is dedicated to digitally storing, analyzing and expressing the spatial-temporal poral information in a family tree. Furthermore, FTGIS aims to construct a visible spatial-temporal network of family tree and to express spatial and genealogical relationships clearly and understandably. In dealing with Chinese family tree sources, mining the driving mechanisms of family inheritance and development to reproduce the history of Chinese civilization is the ultimate goal of the FTGIS. Specifically, the main objectives of the FTGIS are as follows:

(1) To digitize the full texts of family trees. Most family trees are stored in libraries and private homes in the form of printed text, which makes it difficult to analyze and share family tree information. Existing family tree information systems mainly support bibliographic search rather than full-text search. The primary objective of the FTGIS is to digitize the full text of family trees and then build family tree databases and establish a foundation for the construction of the FTGIS platform.

(2) To make the temporal information contained in family trees comparable. The expression of the temporal information is mainly based on the Chinese traditional calendar, supplemented by the Christian era. These two ways of indicating time are different in terms of benchmarks. Hence, the FTGIS is dedicated to

ensuring that the temporal information can be located by using a time conversion engine.

(3) To make the spatial information contained in family trees locatable. Place names are the main spatial information in family trees. These place names are not associated with longitude and latitude coordinates, which makes them difficult to locate. In addition, as time passes, place names, locations, and regions often change. The FTGIS can map ancient place names to specific spatial locations or regions using an encoding technology based on ancient and modern place names.

(4) To build a platform for family tree information collection and sharing. In China, many family trees have not been publicly published, making information collection and sharing an issue. The FTGIS platform aims to provide an efficient and convenient way to collect and share family tree information.

(5) To express the family tree information in a dynamic and visual way. Family tree information is mostly recorded in the form of dry words, which is a disadvantage for intuitively expressing family tree information, especially spatial-temporal information. Expressing family tree information in a dynamic and visual way not only helps the public gain an understanding of family trees but also helps researchers to analyze family tree information. The FTGIS aims to provide various ways of displaying family tree information directly and vividly.

(6) To promote family tree information analysis with the aid of GIS. As a tool, GIS has a variety of functions, including spatial analysis, spatial positioning, and multidimensional visualization. Of these, spatial analysis is the core function. Therefore, expressing and analyzing family tree information by leveraging GIS is both possible and convenient. Through the FTGIS, information on individuals, families and clans can be mined in addition to the path of the development and heritage of any clan.

3.2 A Unified Spatial-Temporal Framework for Family Trees

3.2.1 Why Is a Unified Spatial-Temporal Framework Needed?

A unified spatial-temporal framework is essential for constructing the FTGIS. A large amount of spatial-temporal information is implicated in the preface, personal biographies and other content of a family tree, which is an important part of studying clan lineages, migrations and the spatial distribution of families and building the spatial-temporal pedigrees of families and clans. However, such information is recorded with different spatial-temporal benchmarks. Hence, constructing a unified spatial-temporal framework is extremely important for mapping the spatial-temporal information from different family trees for further analysis. GIS technologies can be used to conduct a spatial analysis of family trees in various time periods and geographic regions.

Time is one basic type of information contained in family trees. Generally, time is expressed in two forms, namely, the Chinese traditional calendar and the Gregorian calendar. These two forms are different in terms of datum. Specifically, the Chinese traditional calendar can be divided into two types in which time is recorded using the annual number of dynasties and the annual branches of dynasties. Ancient Chinese family trees mainly employ the Chinese traditional calendar, while modern family trees mostly use the Gregorian calendar. However, the Chinese traditional calendar usually omits the name and annual number of the dynasty; therefore, that time must be estimated by interpreting the context. It is difficult for a computer to directly compare the different types of temporal information, and issues such as clan lineages, population ages and life regulations cannot be definitively resolved. Thus, constructing a unified temporal datum and unifying the time expression methods are essential.

Spatial information in family trees is mainly expressed by place names and simple maps that lack accurate descriptions of specific locations and regions. Ancient place names are mostly expressed as lower-level place names and tend to omit higher-level administrative divisions. This makes it difficult for modern people to locate ancient place names. In addition, place names change frequently over time. One place may have different names in different time periods, and different places may have the same name. Therefore, locating positions of place names correctly is important for further study of family trees.

3.2.2 How Can a Unified Spatial-Temporal Framework Be Constructed?

Positioning time and place names correctly is the core of building a unified spatial-temporal framework. Therefore, models for time and place names should be built and then combined into a unified spatial-temporal framework. First, to address the time expression issues mentioned above, time must be located based on a unified temporal datum. More importantly, the time model should be able to cover the entire process of the spatial-temporal evolution of Chinese civilization. Second, types of place names should be selected feasibly. Ancient and modern place names are taken into consideration because these two types of place names exist at specific time points or periods. Positioning place names requires building relationships between ancient place names and modern place names and spatially orienting them based on administrative divisions. Then, the current locations of ancient place names can be identified based on this relationship. Moreover, place names can be abstracted as geographic regions or entities with specific spatial locations, shapes, and ranges. In addition, many place names in family trees lack longitude and latitude coordinates. Therefore, a specific spatial-temporal symbol for expressing place name is needed.

A unified spatial-temporal framework for family trees is proposed in this study. This framework uses the Christian era year, Julian date and time as temporal datum and Chinese historical administrative divisions and ancient and modern place names as spatial datum. Through spatial-temporal database technology, the framework converts Chinese traditional time to the Christian era and Julian date and time using a time conversion engine. Thus, time information can be positioned. The framework maps ancient place names to a specific location or extent using an ancient and modern place name encoding engine. Thus, spatial information can also be positioned. Then, the family tree information can be expressed in a unified spatial-temporal framework, and researchers can use spatial-temporal data analysis and mining methods to investigate family tree information.

3.3 FTGIS Data Model

3.3.1 Content and Information of Family Trees

Family trees are rich in content. Generally, a family tree contains a cover, preface, commentary, legend, catalog, compiler, details of origin, lineage chart, Zibei, honor record, biographies, clan rule and domestic discipline, and information about ancestral temples, tombs, clan property, contracts, writings and serial numbers (Ge 1996; Wang 2006). Zibei is a word used in a name to indicate the rank of a clan. Modern family trees usually contain attached demographic charts, compared tables of time, and ancient and modern place name references. Some modern family trees even contain audio and video materials. There is no standard for the content of family trees. Some contain more information, and some contain less; some are brief, and some are detailed.

Family tree information can be divided into three parts: basic information, core information and other information. Time and place are the basic information of a family tree. The birth and death information of family members constitutes the main content of a family tree together with their activities at specific time periods and places. Time and place frequently appear in family trees. Based on a family tree, we can know when and where the ancestor of a branch clan migrated; networks of blood relationships, which imply the order of birth; when and where individuals were born and died; where their graves are located; and when and where they lived, studied, worked and had experiences. Extensive time and place information express the inheritance relationship of a family in every generation from the temporal perspective and convey the distribution and migration information of a family from the spatial perspective.

The core information of the family tree is clan information and individual information and relationships. Clan information includes compiling information, the clan branch and migration. This information shows detailed migration information for a clan and its main migrators, such as the time when a migration occurred and the

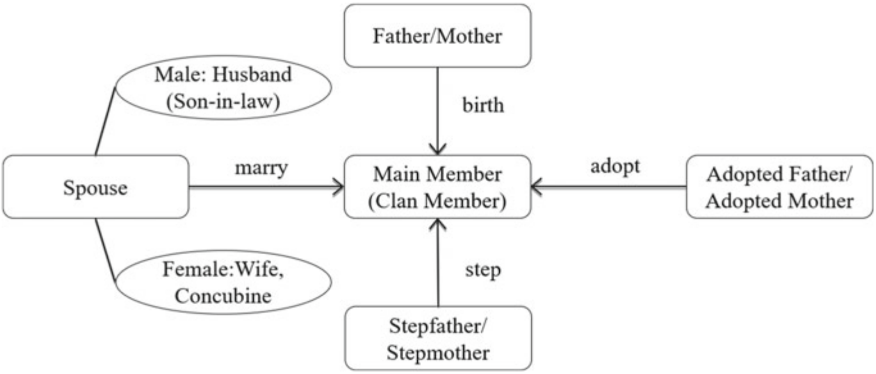

Fig. 3.1 Relationships in a family tree

origin and destination of the clan or migrators. Compiling information includes the preface, commentary, autobiographies, honor records, biographies, stylistic rules, clan rules and collection records. Individual information includes name, Zi, Hao, nation, generation, rank, birth time and place, experiences, death time and place and grave location. Zi and Hao are respectful title for a person used in ancient Chinese society. Experiences include when and where the individual studied, lived and worked. Relationships include family and clan relationships. As illustrated in Fig. 3.1, these relationships include father, mother, spouse, stepfather, stepmother, adopted father and mother, and main member or clan member.

Other family tree information, also called bibliographic information, includes genealogy place, genealogy name, compiler, compilation mode, version, carrier form, binding form, annotation, abstract and collection unit.

3.3.2 Overview of the Models

Based on the above analysis, this study proposes the FTGIS data model, which is composed of five data models. Figure 3.2 shows the components of the FTGIS data model and their relationships. These five data models can be divided into two categories: the spatial-temporal framework data model and the family tree spatial-temporal data model. The time data model and place name data model constitute the data model for the unified spatial-temporal framework. The family tree spatial-temporal data model includes the family tree bibliographic model, family tree item content model and family tree lineage record model.

Figure 3.3 shows the time data model, which is divided into two parts. One part contains the entities HistoricalStage, Dynasty, DynastyStage, Emperor and EmperorReignTitle. The other part is the time reference, including YearRef and DateRef entities. The HistoricalStage entity contains the id, name, start and end date attributes. The Dynasty entity contains the id, id of HistoricalStage entity, name, start and end