WENWEN LI
School of Geographical Sciences and Urban Planning
Arizona State University
Email: wenwen@asu.edu

Wenwen Li is Associate Professor of Computational Spatial Sciences at Arizona State University. She is the founding director of the Cyberinfrastructure and Computational Intelligence Lab (CICI, http://cici.lab.asu.edu), and Associate Director of SPatial Analysis Research Center (SPARC). Her main research interest is to develop new theories, methods and tools to advance spatial data retrieval, real-time visualization of big data, as well as high- performance spatial analytics. Li's research has been widely applied to solving problems in both physical and social science domains, such as polar climate change, disaster management, terrain analysis, urban heat island etc. She has published over 100 papers in top-notch journals and conference proceedings and has served as the principle investigator (PI) or co-PI for multiple NSF projects in cyberinfrastructure and GeoAI (geospatial artificial intelligence). Li was the recipient of the 2015 NSF CAREER award, the most prestigious award of NSF to young scientists.

Position Statement

The proliferation of geospatial big data has brought ample opportunities for the scientific community towards better understanding and solutions for major problems our society is facing—environmental and climate change, pollution, traffic congestion, disaster, urban heating, rapid communications of infectious diseases, to name a few. However, the ever- exploding geospatial data, the diversity in data source, heterogeneity in data formats, the high velocity at which they are being produced create significant challenges in the storage, retrieval, processing and visualization of such data. Traditional GIS algorithms and software has to be renovated to accommodate this paradigm shift towards data-intensive research and data- driven science.

In its 2003 blue-ribbon report, NSF defined cyberinfrastructure as the next generation research infrastructure that relies on high performance computing, high-speed internet and advanced middleware to solve data- and compute- intensive problems at a scale that is not possible before (Atkins 2003). This report highlights four key research areas of cyberinfrastructure, which in my view can contribute to the development of the research core for the emerging spatial data science. These four areas are: high performance computing, data access, visualization, and virtual organization. Geospatial big data, has injected new ingredients for deepening and spatializing these research topics. A growing literature can be seen in the past decade or so, however, many interesting and unique research questions related to geospatial data are yet to be answered. For instance, geospatial data is known to be heterogeneous – they could be structured or unstructured, they can be encoded in image formats or plain text, and they may be various in quality due to the collection methods.

Hence, paradigm for geospatial data retrieval would be very different from information search such as what Google is doing (Li 2018). How to make effective discovery and search of geospatial data at a user's space and time of interest, how to perform accurate data retrieval in the absence of good metadata or using the metadata that are semantically ambiguous? How to leverage spatial-visual analytics to easily refine research results and quickly narrow down to the subset of data which is most needed by end users? How to deliver spatial methods, tools, models in the same way as we perform data search?

Another important aspect is high-performance computing, which may be the most researched topic in cyberinfrastructure. Many existing solutions, including CyberGIS (Wang 2010), adopt divide-and-conquer strategy to enable scalable data processing—namely to process massive amount of data utilizing distributed computing and batch processing. These applications cannot satisfy the need for real-time processing, which not only requires processing large data, but also streaming data that arrive in real-time. Data of the latter kind can be increasingly found as more smart cities IoT devices are being deployed and functioning. Streaming architecture has emerged to become the state-of-the-art to achieve both high throughput and low-latency in data processing. However, very limited research has done in the geospatial domain. How to perform real-time spatial indexing on the streaming data? How would the data storage model change to accommodate such high-velocity data? How to make in-time spatial decisions based on these data?

What is missing from Cyberinfrastructure research is the intelligent analytical core for future GIS. CyberGIS is deemed as the integrated framework of cyberinfrastructure, GIS and spatial analysis, in which spatial analysis and spatial statistics are the mainstream methods for spatial data processing. However these methods are better at handling "small data" with fewer errors, and the distribution of the data has to follow some predefined assumption. The veracity of the analytical results may be challenged when the data follows a different distribution than expected. The scale at which the best results can be obtained also needs iterative experiments. In the big data world today, theory and principles underlying these data are not fully discovered yet. Hence, there is an urgent need of a data-driven, theory-free method that can handle massive data and extract useful patterns from them. GeoAI, or geospatial artificial intelligence, a transdisciplinary extension of AI, is seen as a promising research area that contributes to building the smart brain for future GIS and for setting the agenda for spatial data science. As an emerging field, many interesting questions can be asked. How to inject spatial principles (spatial heterogeneity and spatial autocorrelation) into the deep machine learning (comparing to shallow models mentioned above) to enable a spatialized learning paradigm? How to address the issue of lack of training data in the geospatial domain? How to increase the interpretability in machine inference process? The author has recently started a NSF project working on these questions and she looks forward to the communication with other symposium participants to move this research area forward.

References:

Atkins, D. (2003). *Revolutionizing science and engineering through cyberinfrastructure: Report of the National Science Foundation blue-ribbon advisory panel on cyberinfrastructure*.

Li, W. (2018). Lowering the barriers for accessing distributed geospatial big data to advance spatial data science: the PolarHub solution. *Annals of the American Association of Geographers 108*(3): 773–793.

Wang, S. (2010). A CyberGIS framework for the synthesis of cyberinfrastructure, GIS, and spatial analysis. *Annals of the Association of American Geographers 100*(3): 535–557.