ATSUSHI NARA
Center for Human Dynamics in the Mobile Age
Department of Geography
San Diego State University
Email: anara@sdsu.edu

Atsushi Nara is an Associate Professor in the Department of Geography and Associate Director for the Center for Human Dynamics in the Mobile Age at the San Diego State University. He received his Ph.D. in geography from the Arizona State University. He worked as a post-doctoral research scientist at the Center for Spatial Analysis at the University of Oklahoma.

Nara's research interests focus on Geographic Information Science, spatiotemporal data analytics, geo-computation approaches, agent-based modeling and complex adaptive systems, applied to study human mobility, urban dynamics, and interdisciplinary fields. In recent years, his research works has been supported by National Science Foundation, National Institute of Health, and California Air Resource Board to apply GIScience and geo-computation for: supporting spatial decisions under emergency evacuation; identifying strategies to build capacity for geo-computational thinking in preK-14 education; examining spatial disparities in public health and cancer epidemiology; and understanding the impacts of air pollution in the California-Mexico border region.

Challenges in Spatial Data Science and GeoComputation for Human Dynamic Research

The increasing adaptation and usage of location aware mobile technologies, sensor devices, and ubiquitous cyberinfrastructure in everyday life generate a massive amount of data about space, place, and human activity dynamically over time. Open data initiatives and citizen-based sensing via web/mobile/social media platforms further enable access to a large amount of fine-granular and individual-scale data, which were not available in the past. This provides opportunities for researchers to explore and discover spatial and spatiotemporal knowledge about human dynamics that is buried in very large, high-dimensional, and complex datasets. Human dynamics is a transdisciplinary research field that focuses on the understanding of dynamic patterns, relationships, narratives, changes, and transitions of human activities, behaviors, communications, and environments. Human dynamics research benefits from the availability of big geospatial data leading to a data-driven geospatial scientific, purely inductive and emergent forms of analysis that data to speak for itself (Kitchin, 2014; Kwan, 2016). With the technological advancements and big geospatial data availability, researchers can now trace, monitor, map, analyze and model the spread of human and social movements, disease outbreaks, nature hazards, crime incidents, and popular events; however, there exist notable challenges including but not limited to issues with high-performance computing, large geospatial data management, biases, privacy and ethics, data and algorithm uncertainties, heterogeneous data integration, and education (Evans et al., 2019; Kwan, 2016; Nara, Tsou, Yang, & Huang, 2018; Tsou, 2015). This position paper highlights four challenges to move forward with human dynamics research utilizing new technologies and big geospatial data.

One of key challenges is related to data uncertainty. In spite of the emerging new research opportunities to produce spatial and spatiotemporal knowledge by utilizing big geospatial data, most of these data are not the output of instruments designed to produce valid and reliable data amenable for scientific analysis (Lazer, Kennedy, King, & Vespignani, 2014). For example, the quality of location information in social media data can be controlled by end users, which makes researchers challenging to know the level of uncertainty. In case of Instagram, a location of an Instagram post is selected by a user based on a list of nearby point of interests (POIs) provided by Instagram; therefore, a user can easily manipulate his/her location. There also exist quite a few web tools and mobile applications to fake location information. While a user may decide to fake (or spoof) location to protect individual's geo-privacy, few studies have discussed and incorporated location spoofing in the existing GIScience literature (Zhao & Sui, 2017). Our recent research project also discovered that emission source data published from a governmental agency in Mexico include data with large locational inaccuracies.

Related to data uncertainty, algorithm uncertainty is another challenge. Kwan (2016) questioned that big data-driven research ignores the potentially significant influence of algorithms on research results, and thus geographic knowledge generated with big data might be more of an artifact of the algorithms used than the data itself. For example, Fischer (2014) mapped six billion geo-tagged tweets and observed a banding phenomenon, where the original tweet locations tend to align with the closest latitude or longitude, suggesting that tweet locations might have been fuzzed by Twitter through snapping them to the closest latitude or longitude to prevent people's exact locations being disclosed. Researchers often do not have access to, or even do not know about such algorithms being used by data providers who generate, process, and provide their data through Application Programming Interfaces (APIs). Moreover, in order to deal with big data, algorithms are increasingly implemented as computerized procedures, and they become increasingly detached from and less visible to researchers who use them (Kwan, 2016). Consequently, such algorithms introduce greater uncertainty and potentially result in significant differences in research findings. Hence, it is crucial to examine and evaluate the validity of data and algorithms in order for maximizing the utility of big geospatial data.

The integration of heterogeneous data from multiple sources is a way of producing new insights. In particular, quantitative and qualitative data integration plays a key role to enrich spatial data analytics. In the era of big data, a wide variety of data is available from public and private sector organizations as well as citizens via internet and social media. However, data analytics that focuses only on numerics and algorithm are not enough to study 360-degree of human activity or relationship that includes unforeseeable factors (Shacklett, 2015). New data analytics frameworks and methodologies are needed to solve data integration issues such as scales (e.g., big vs small), formats (e.g., relational vs object-based), types (e.g., quantitative vs qualitative), and biases.

Last but not least, with the increasing demand for spatial computational (or geocomputational) skills in spatial data science research and industry, the development of effective learning pathways toward geocomputationally intensive jobs is prominent (Dony et al., 2019). This new demand requires interdisciplinary domain knowledge in spatial statistics and analysis, geocomputational skills (or hacking skills), and substantive expertise (Figure 1). Two key challenges on spatial data science and geocomputation education are (1) the need for educational cross collaborations between Geography and Computer Science to cultivate spatial and computational thinking; and (2) inclusiveness and diversity to broaden participation in the context of computational geography.

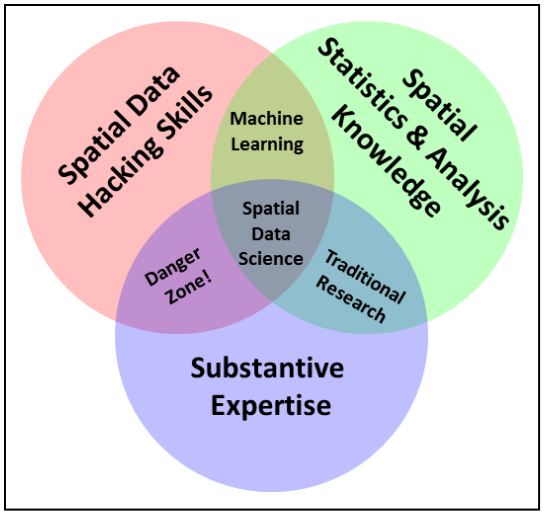

Figure 1. Spatial data science Venn diagram modified from Conway's data science Venn diagram (Conway, 2010).

References

Conway, D. (2010). The Data Science Venn Diagram. Retrieved September 27, 2019, from Drew Conway website: http://drewconway.com/zia/2013/3/26/the-data-science-venn-diagram

Dony, C., Magdy, A., Rey, S., Nara, A., Herman, T., & Solem, M. (2019, February 27). RPP for geocomputation: Partnering on curriculum in geography and computer science. *RESPECT (Research on Equity & Sustained Participation in Engineering, Computing, & Technology) 2019*, Minneapolis, MN. (Poster)

Evans, M. R., Oliver, D., Yang, K., Zhou, X., Ali, R. Y., & Shekhar, S. (2019). Enabling Spatial Big Data via CyberGIS: Challenges and Opportunities. *CyberGIS for Geospatial Discovery and Innovation*, 143–170. https://doi.org/10.1007/978-94-024-1531-5_8

Fischer, E. (2014). Making the most detailed tweet map ever. Retrieved October 15, 2016, from Mapbox website: https://www.mapbox.com/blog/twitter-map-every-tweet/

Kitchin, R. (2014). Big Data, new epistemologies and paradigm shifts. *Big Data & Society*, *1*(1), 2053951714528481. https://doi.org/10.1177/2053951714528481

Kwan, M.-P. (2016). Algorithmic Geographies: Big Data, Algorithmic Uncertainty, and the Production of Geographic Knowledge. *Annals of the American Association of Geographers*, *106*(2): 274–282. https://doi.org/10.1080/00045608.2015.1117937

Lazer, D., Kennedy, R., King, G., & Vespignani, A. (2014). The Parable of Google Flu: Traps in Big Data Analysis. *Science*, *343*(6176): 1203–1205. https://doi.org/10.1126/science.1248506

Nara, A., Tsou, M.-H., Yang, J.-A., & Huang, C.-C. (2018). The Opportunities and Challenges with Social Media and Big Data for Research in Human Dynamics. In S.-L. Shaw & D. Sui (Eds.), *Human Dynamics Research in Smart and Connected Communities* (pp. 223–234). https://doi.org/10.1007/978-3-319-73247-3_12

Shacklett, M. (2015). Thick data closes the gaps in big data analytics. Retrieved October 10, 2019, from TechRepublic website: https://www.techrepublic.com/article/thick-data-closes-the-gaps-in-big-data-analytics/

Tsou, M.-H. (2015). Research challenges and opportunities in mapping social media and Big Data. *Cartography and Geographic Information Science*, *42*(sup1), 70–74. https://doi.org/10.1080/15230406.2015.1059251

Zhao, B., & Sui, D. Z. (2017). True lies in geospatial big data: Detecting location spoofing in social media. *Annals of GIS*, *23*(1): 1–14. https://doi.org/10.1080/19475683.2017.1280536