CLIO ANDRIS
School of City + Regional Planning
School of Interactiv Computing
Georgia Tech
Email: clio.andris@gmail.com

Clio Andris is an Assistant Professor with a joint appointment in the School of City + Regional Planning and School of Interactive Computing at Georgia Tech. She researches spatial social networks, social flow data, interpersonal relationships, social life, GIS, and urban planning.

Her PhD is in Urban Information Systems (Dept. of Urban Studies and Planning at MIT, 2011), where she was a National Defense Science and Engineering Graduate (NDSEG) Fellow and a member of the MIT Sense*able* City Lab. She has an MS in Geography (University of South Carolina), a BA in American Studies (Boston University), and held postdoctoral positions at the MIT-Singapore Alliance for Research and Technology and at the Santa Fe Institute, studying complex systems.

Andris' background and expertise is in Geographic Information Science, innovating in this field by integrating evidence of social relationships into geographic models and representations of geographic space. She is forging a new interdisciplinary topic of study, Interpersonal Relationships in Geographic Space, which has spatial analytic, statistical, modeling, geo-visualization, qualitative, quantitative, case-study and storytelling components. Her research is conducted by using mixed-methods approaches, and has distilled her research program into data-driven analyses and method development for (1) geolocated social networks, (2) measurements of geographic social capital, (3) geolocated dyadic relationships and (4) large-scale social flow data. Her research focuses on the quality of information, data structures and types, big data, how to quantify behavior and relationships, and how to use relationship-enriched GIS models for urban planning, transportation and civil engineering problems.

Andris serves as director of the Friendly Cities Lab, founded in 2015, to promote "data-driven love for community." Having worked with 20+ students, the lab has published a number of works in *Transactions in GIS, Computers, Environment and Urban Systems, EPJ Data Science PLOS ONE, International Journal of Geographical Information Science, Environment and Planning A, The Professional Geographer*, and in peer-reviewed conference proceedings in the ACM and IEEE. They use creative datasets from AirBNB, Yelp, the Yellow Pages, Big Brothers Big Sisters, and NCAA, as well as call data records (CDRs), i.e., mobile phone data, through projects with AT&T, SingTel, British Telecom, Orange and most recently, Turk Telecom.

On Setting the Spatial Data Science Agenda

DEFINE SPATIAL DATA SCIENCE—We should define what we mean by spatial data science. Data science has emerged as a degree program and field of study—not to mention as the precursor to wide calls for "data scientists" in the professional hiring world. But is there really a science to it, with a straightforward method, and aim to simply test the properties of something? Or is it more engineering—*getting things to work*? If science lends itself to examining and generating laws, and engineering to building and performing a task, what is the role of data science in the modern technical world? What will the purview be, how is it different than geographic information science? What is the next big term to emerge?

HUMAN BEHAVIOR—Because much of our data is generated by individuals, we should continue to connect with studies of "human behaviour," and improve this connection by identifying research that can enrich our research questions and data analyses. With the integration of new datasets from mobile units, social media, human consumption records, telecommunications, migration and other kinds of flows, new challenges arise. Relatively straightforward challenges (*At what scale should I perform my analysis?)* give way to more nuanced questions that require input from other fields of research: *Why are people performing this action? What are the implications? How is it helping them reach a certain goal? What does it mean to do this? How can the space/place be configured to best suit the needs of the people?* Human behavior differs from demographics in so many ways, and to do richer, more meaningful analysis, we must (continue to) meaningfully engage with communities in psychology, communications, sociology, etc. to explain why we are seeing certain results.

DATA PRODUCTS—Currently, spatial data is hosted on the web and in proprietary locations, and data fusion continues to be a research challenge. Pre-packaged data that combines multiple sources could help user communities apply spatial data into their decision-making and analyses. For example, census data on mobility and travel patterns can be augmented with newer data types from social media and GPS traces but it is hard to know which to use, and if a user doesn't have the skills to clean and manage the larger datasets, they may miss out on the richness of new data sources. A successful example of a data product is the Social Vulnerability Index, created by Susan Cutter and team in the early 2000s. This index is available at the census division level and incorporates many different input factors (elderly population, children, minority population, low-income population, etc.) and has been adopted by large agencies such as FEMA as a go-to data set for determining how natural disasters may impact different communities in different ways. These types of indexes can be dangerous because their input variables (and weights) can be subjective—including confounding variables and neglecting important variables—to create a bias. However, when done carefully, these multi-source datasets can save agencies time and effort, and as a result, the types of input variables that are important to our expert community can be used as factors in decision-making. These may include data on accessibility, human mobility, vulnerability, social capital, digital literacy, human agency, loneliness, stress, mental illness, general deprivation, at-risk populations, etc.

IMPACT—When a new method is published, there is often far too little proof that it could help solve an actual problem or give new insights to users. Spatial data science and especially urban analytics should make it a standard and a norm for authors to illustrate how their new method adds new information. As of now, most research in spatial data science scantly reflects on the location. These papers tend to want to contribute to advancing methods, but they sidestep any meaningful discussion about their case study. Many papers in spatial data science use a dataset to illustrate their method. This spatial dataset (ex. e-scooters in New York, tweets in Rome, taxis in Jakarta) has a location and a population—it does not exist in a vacuum. Our papers need to describe the case study better, and make recommendations based on their results. We can encourage researchers to go beyond a brief statement of “this method can be used for policy makers, urban planners and geographers” at the end of a paper, but to step the reader through how this could be used. This exercise may be uncomfortable for some spatial data analysts, as it may require external reading and/or communications with practitioners, but it should be the responsibility of the analyst to draw a connection between their work and how it could be used in the “real world.” Too many urban planners, civil engineers, and policy makers are not using any advanced spatial data mining methods or analysis to make decisions, in part because spatial data analysts simply have not made journal articles and their content accessible to busy, everyday workers. Could we get academia to reward these kinds of efforts? Could we ensure that “applied” work is not seen as a “lesser” endeavor? Do we need better role models to set a higher standard for working with others to implement our techniques—or do we not have the power to do so?

CS TECHNIQUES—We should continue to connect with high-performance computing and cutting-edge computer science techniques in the Natural Language Processing, Image Segmentation, Social Computing, Augmented Reality, Human-Computer Interaction, Artificial Intelligence, Machine Learning/Deep Learning research communities. Continue to integrate these cutting-edge techniques into spatial research. Keep abreast of the latest technologies and methods for data analysis and apply them to spatial problems. Teach students programming techniques as well as high-quality mapping techniques.

NETWORK—We should identify graduate programs, journals and conferences that are core to the beliefs and ethics of performing good spatial data science. Identify labs and research groups in both academia, private sector, NGOs and government organizations that engage with the mission of spatial data science and are willing to adhere to the core values of what spatial data scientists believe (on ethics, privacy, etc.) and what their mission is.

DATA STANDARDS—We should create better ways to evaluate our data sets. How do we know if data is reliable or not? One could argue that the government is now not the leading provider of spatial data—changing the landscape of the responsibility of private sector data providers (Zillow, Visa, Twitter, etc.) to data users. We should envision the role of government vs. industry in creating data sets and argue for better data standards. We can also ask: How can we persuade private sector companies to share their data with the public? Do data sources have an ethical responsibility to disclose crucial information about their data (collection hiccups, added noise or imputed information)?

INCLUSION—The field of spatial data science is a privileged field that tends to engage students with good opportunities for education. Spatial data science can work on committing to diversity by including underrepresented voices and giving these communities a chance to be part of the data collection, analysis and impact/reflection processes. In addition, spatial data science should continue to commit to open source technologies and free, accessible data.