Integers: Interpretable Time Series Autoregression for Periodicity Quantification

Xinyu Chen (Postdoctoral Associate), Department of Urban Studies and Planning, MIT
GitHub repository: `https://github.com/xinychen/integers`
Role: Creator & Primary Developer
Contact information: `xinychen@mit.edu`

The "**integers**" project focuses on discovering the hidden rhythms of urban mobility, climate systems, and user behaviors in digital platforms by understanding patterns in how real-world systems change over time. From urban mobility and public transit ridership to climate variables and web traffic (see Figure 1), many real-world systems exhibit predictable patterns, known as *periodicity*, on multiple overlapping cycles–daily, weekly, seasonal, and annual. Human mobility in urban areas often shows recurring patterns like morning and afternoon rush hours. These rhythms are fundamental to how cities function, influencing transportation planning, transit schedules, and even emergence response. Similarly, climate systems exhibit strong periodic patterns, such as yearly seasonality in temperature and precipitation. These patterns evolve over time due to long-term variability and climate change. To discover spatiotemporal periodic patterns that are easily understandable and actionable, this project proposes a novel, interpretable machine learning framework called "multidimensional sparse autoregression". Open datasets such as long-term multi-modal human mobility data in New York City (NYC) and Chicago, multi-decade global climate variables, and large-scale Wikipedia page views enable a shift from purely analytical studies to data-driven experiments.
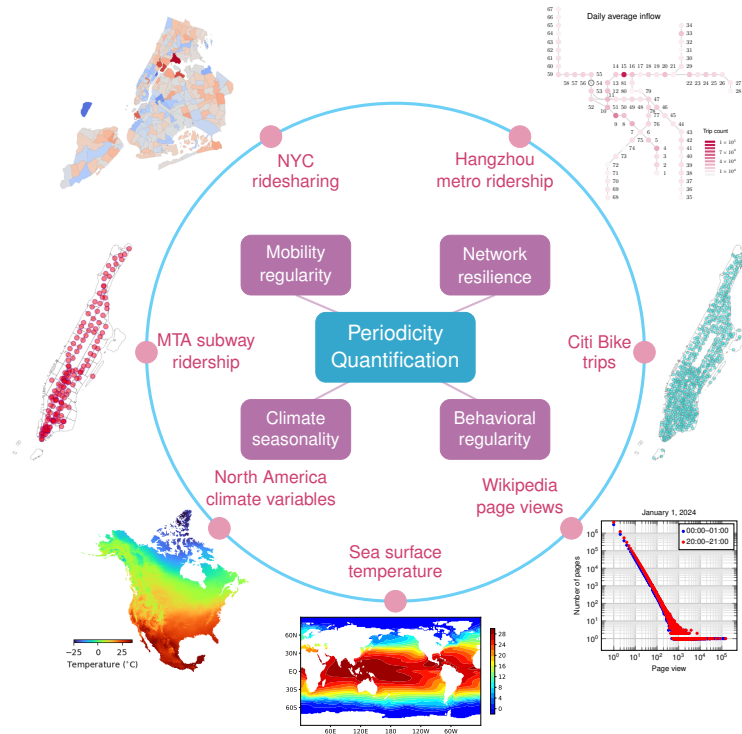

Figure 1: Conceptual overview of the diverse open datasets for periodicity quantification.

data alignment of Wikipedia page view observations and utilize the heavy-tail distribution to select ...

This project made substantial efforts to improve the data quality, alignment, and reproducibility in Python programming. First, this project aggregates billions of trip records of ridesharing, taxi, subway, and bikesharing in NYC and Chicago as multidimensional arrays in the format of compressed NumPy array (i.e., .npz) across several years. In particular, the trip records of subway and bikesharing are usually associated with station information, which are different from other transportation modes, thus the subway and bikesharing data are projected onto taxi/ridesharing areas (see e.g., Figure 2 in Manhattan of NYC).

The interdisciplinary insights generated by this project have significant impacts, including 1) understanding urban dynamics and policymaking, 2) assessing disruptions and recovery of the COVID-19 pandemic and informing strategies for urban resilience, 3) multi-modal transportation planning and transit scheduling, 4) climate science and environmental resilience associated with temperature and precipitation variables, 5) measuring behavioral regularity in digital platforms, and 6) advancing interpretable machine learning. The project also highlights the importance of creatively utilizing open data and fostering transparency and reproducibility. While designed for large datasets, the emphasis on interpretable, quantifiable periodicity implies that this framework can empower "small data" projects in disciplines where open data practices are nascent.
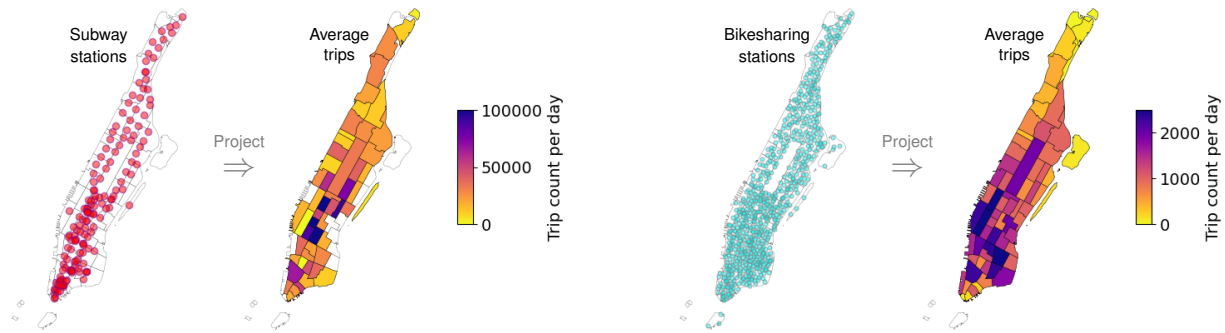

Figure 2: **(Left)** Subway stations are projected onto 52 areas in Manhattan (i.e., 69 areas in total). **(Right)** Bikesharing stations are projected onto 67 areas.

Figure 3: The last mile of reproducible research for periodicity quantification.

The work related to this project includes:

- **Xinyu Chen**, Vassilis Digalakis Jr, Lijun Ding, Dingyi Zhuang, Jinhua Zhao (2025). Interpretable time series autoregression for periodicity quantification. arXiv:2506.22895.
  `https://arxiv.org/abs/2506.22895`
- **Xinyu Chen**, Qi Wang, Yunhan Zheng, Nina Cao, HanQin Cai, Jinhua Zhao (2025). Data-driven discovery of mobility periodicity for understanding urban systems. arXiv:2508.03747.
  `http://arxiv.org/abs/2508.03747`