**Problem Context**

Basic research questions in analysis of movement behavior pertain to identifying popular places and the routes that people use to travel between them. Patterns and cluster analyses are typically applied in these scenarios to find areas that many people pass, those that many frequent, or routes traveled by many people. Outcomes from these analyses characterize human mobility, such as commuting or recreation, and may illuminate the processes which motivate those behaviors. Understanding the processes that drive mobility is key in generating policies and planning that can effectively manage movement behavior. Considering that many new movement data sets contain higher temporal resolution (e.g., sampling a GPS point every 30 seconds or less) than those previously available, there are opportunities to examine behavior at finer scales that might reveal new information about mobility patterns. Currently, most analytic methods rely on simplified trajectories that represent idealized paths with removed uncertainty.

New, high resolution movement data sets are often generated from Big Data sources. The term “Big Data” refers to information that is collected frequently, with great volume, from a wide variety of sources. The magnitude of these new spatial data presents both challenges and opportunities in GIScience. Spatial data are available at increasingly fine spatial and temporal resolutions, but current methods may not be appropriate for handling the volume or resolution of Big Data. When questions about the utility of Big Data remain in the absence of methods that can handle it. The issue of scale is commonly discussed in geographical contexts as well, though its influence is commonly ignored and unquestioned in movement analysis.

There is some evidence that method selection and variation in temporal resolution may affect results in movement analysis and that over or under-sampling in GPS trajectories has implications for the influence of error. For example, a moving object’s path could be represented as a series of many points (i.e., XY locations taken at some regular time interval) or it could be abstracted to just two points representing the origin and destination of the trip (Figure 1). Between differences among analytical methods and the difference in data representation, it is possible that conclusions about movement behavior for the same trip would differ. The implications of varied temporal resolutions on the conclusions we make about movement behavior on networks remain unclear in the absence of systematic analysis.

Movement, Big Data, and Bicycling

Bicycling is one type of movement behavior that may benefit from knowledge discovery associated with pattern analysis outcomes, though little research has given detailed information about where, when, and how people use bicycles. A range of individual and population-level benefits are associated with bicycling. Despite benefits, levels of bicycling for commute and recreation remain low among United States population. There is a need for research to generate detailed information about where and when people use bicycles so the processes that motivate and encourage riding can be better understood. Traditional methods for generating information about bicycle use are problematic, and Big Data may help fill in some of our current knowledge gaps. Though prior research has demonstrated that data from bicycle share programs may be used to discover commute patterns and characteristics, there have yet to be any systematic studies comparing the activity spaces represented by traditional data to bicycle share data or bicycle trip tracking websites and apps in the same region.

**Research Questions**

The proposed research aims to provide strategic insight into potential changes in conclusions about bicycling movement patterns as varied data sources and temporal scales are considered. The specific research questions addressed are:

1. How do the activity spaces generated from different sources of cycling data vary?
2. How do our conclusions about cycling movement behavior change when using varied data sources and varied levels of temporal resolution?
3. In what ways do these varied data sources indicate which bicycling movement data are not captured, and how can they be combined to advance our understanding of cycling movement processes?

**Data and Methods**

Data gathered from three sources in the Phoenix metro area will be utilized to examine bicycling movement behavior. The first two sources, the Grid Bike bicycle share program and the Strava cycling app, represent Big Data sources (Figure 2). The third source is bicycle count data from selected intersections in the area represents traditional bicycle use volume data.

Initial exploration of the data will focus on geographic data mining to inform knowledge discovery. The aim of geographic data mining is to extract information about processes within certain area, i.e., insights gleaned can then be used as knowledge to inform existing or other information about the same area or process could be integrated to discover further insights. In a movement context, data mining seeks to detect associations between phenomena, or patterns. Mining for the current application domain will indicate differences in movement behavior and activity spaces between data sources. Mining approaches will include measures of similarity such as lengths, temporal patterns, masses, and weights. Clustering methods will also be applied. Pattern and cluster methods examine space, time, or space–time among movement data to discover broader relationships that might exist. While patterns may not match exactly, a pattern or number of locations clustered in a particular area may represent a place or phenomenon of interest.

**Significance**

This research will begin to answer questions about the utility of Big Data that remain in the absence of systematic analysis. Studying how conclusions about movement behaviors (i.e., results from pattern and cluster analysis) change as a result of the data source and varied temporal resolution will facilitate meaningful future investigation by establishing thresholds for data suitability and collection in relation to specific research questions regarding movement. Comparison between the activity spaces generated from varied bicycling data sources will delineate the activity captured through each data collection method. These results will help determine whether these different types of data have common characteristics and therefore, whether Big Data that captures bicycle activity can or should be used to augment traditional bicycling data. In the same vein, results will guide future research design by indicating avenues for connecting with bicycle users so that supplemental data may be collected. A better sense of the routes bicyclists take would better inform improvements that might encourage or facilitate cycling. This knowledge could be used to guide planning decisions and infrastructure improvements by examining bicycling activity in regions where there is interest in encouraging bicycle use.

**References**


