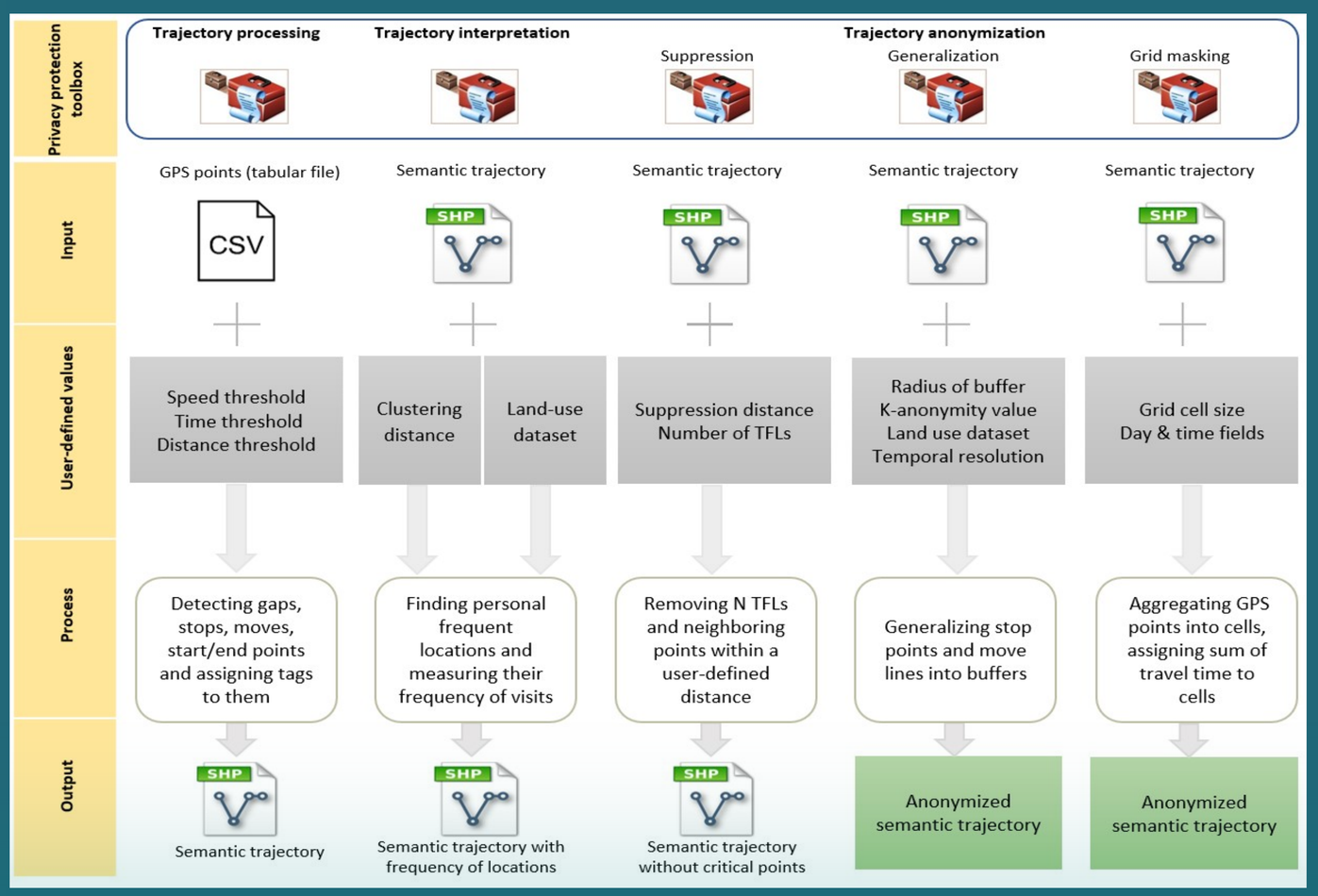
An ArcGIS toolbox for Semantic Trajectory Construction and Privacy protection

Roya Shourouni

Department of Geography, Simon Fraser University, Burnaby, BC, Canada

OBJECTIVE

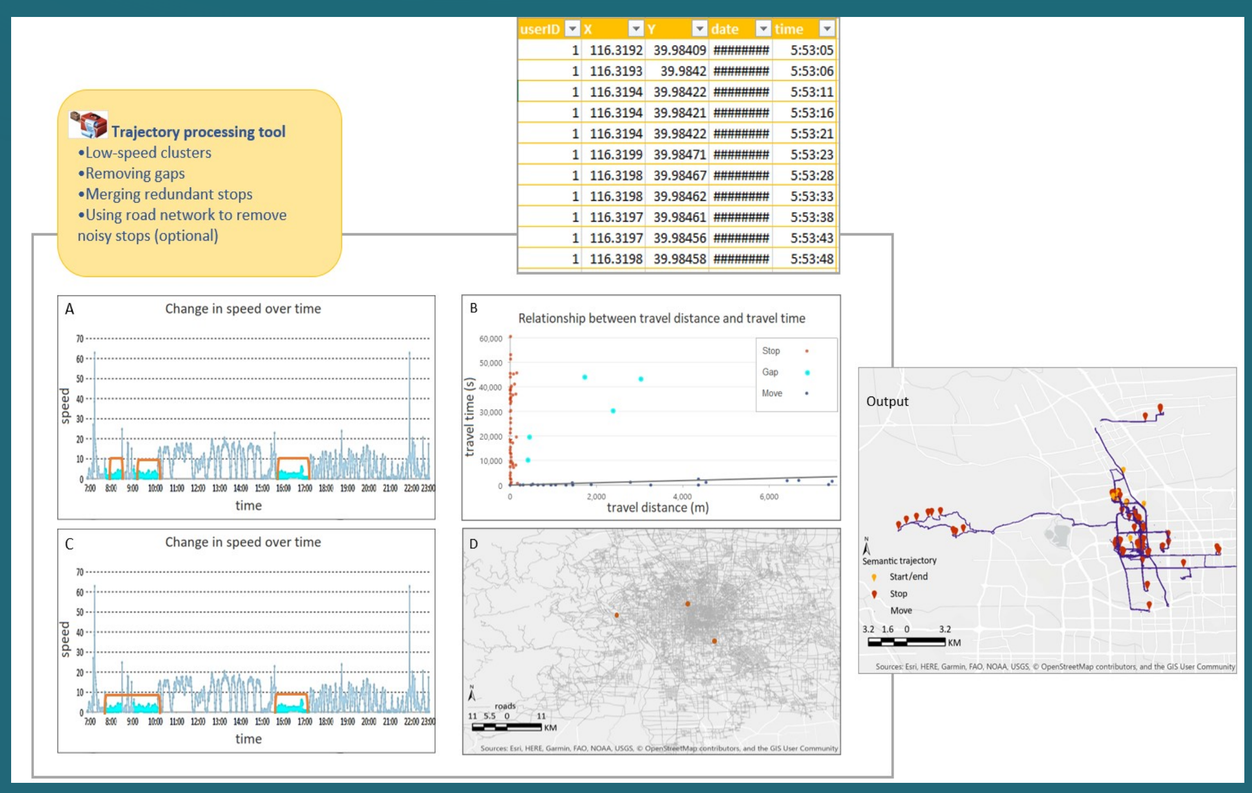
Recently, monitoring the spread of covid-19 has been an emergent task through collecting their movements. However, due to inherent vulnerability of individuals' movements, AKA trajectories, publishing this data combined with external information leads to potential privacy breach. To protect privacy, an anonymization process, which transforms data to remove sensitive information, can be performed but at the cost of a decrease of data utility. Generating tools to ease this process is crucial but poorly perceive d in the academic geography. In this paper, we developed a privacy protection python toolbox in an ArcGIS environment with easy-to-use interface, that provides several tools from rajectory data re-construction to anonymization to novice GIS users. The main objective of this toolbox is to construct semantic trajectories from raw GPS points based on user-defined values, anonymize different parts of semantic trajectories, stops and moves, using the combination of suppression with generalization or grid masking techniques, and then evaluate the resulting balance between privacy and utility by adjusting anonymization parameters. It executes each of this process on thousands of points in seconds. This toolbox offers two advantages to users: 1) users can download and use the easy -to-use toolbox and graphic user interface and comprehensive tooltips without deep knowledge of anonymization methods, 2) due to being accessible and re-usable, it enables researchers in the area of privacy protection to review, reproduce and compare their empirical works.

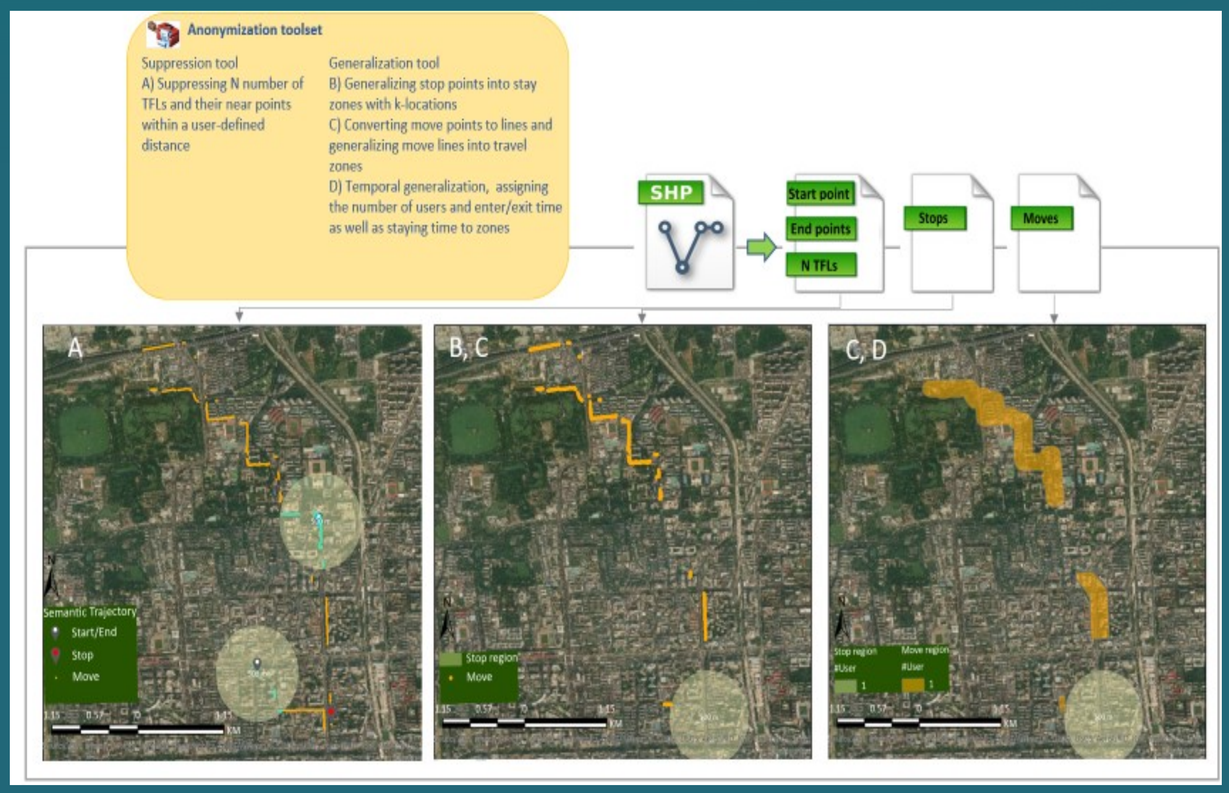


ARCHITECTURE

Using ArcGIS to perform spatial data processing, interpretation and privacy protection offer challenges for the average user since it requires knowledge of coding and an extended workflow as shown in figure below that may be time-consuming, prone to errors and difficult to implement. The toolbox is subdivided into three toolsets and five tools. The first toolset processes raw trajectory data and extracts the meaningful structure of daily trips that delineate the start/end, stops and moves using a stop detection method and user -defined parameter values by means. Next, Trajectory interpretation toolset, consists of two tools, to identify personal frequent locations and shift them to a single meaningful location while measuring the frequency of their occurrence. One tool spatially clusters personal start/end/stops using mean-shift and another tool corresponds personal them to an activity location in a land-use dataset while measuring their frequency of visits. Last, trajectory anonymization, toolset comprises of suppression, generalization, and grid masking tools. Suppression, removes a number of frequent locations as well as their neighbouring points within a user-defined distance. Generalization coarsens the spatio-temporal resolution of the dataset by replacing exact timestamp by approximate time interval and replacing stop points by larger stay zones with k-locations and move points by larger travel zones. Alternatively grid masking tool, blurs GPS points into grid cells while assigning accumulative visit time to each cell.

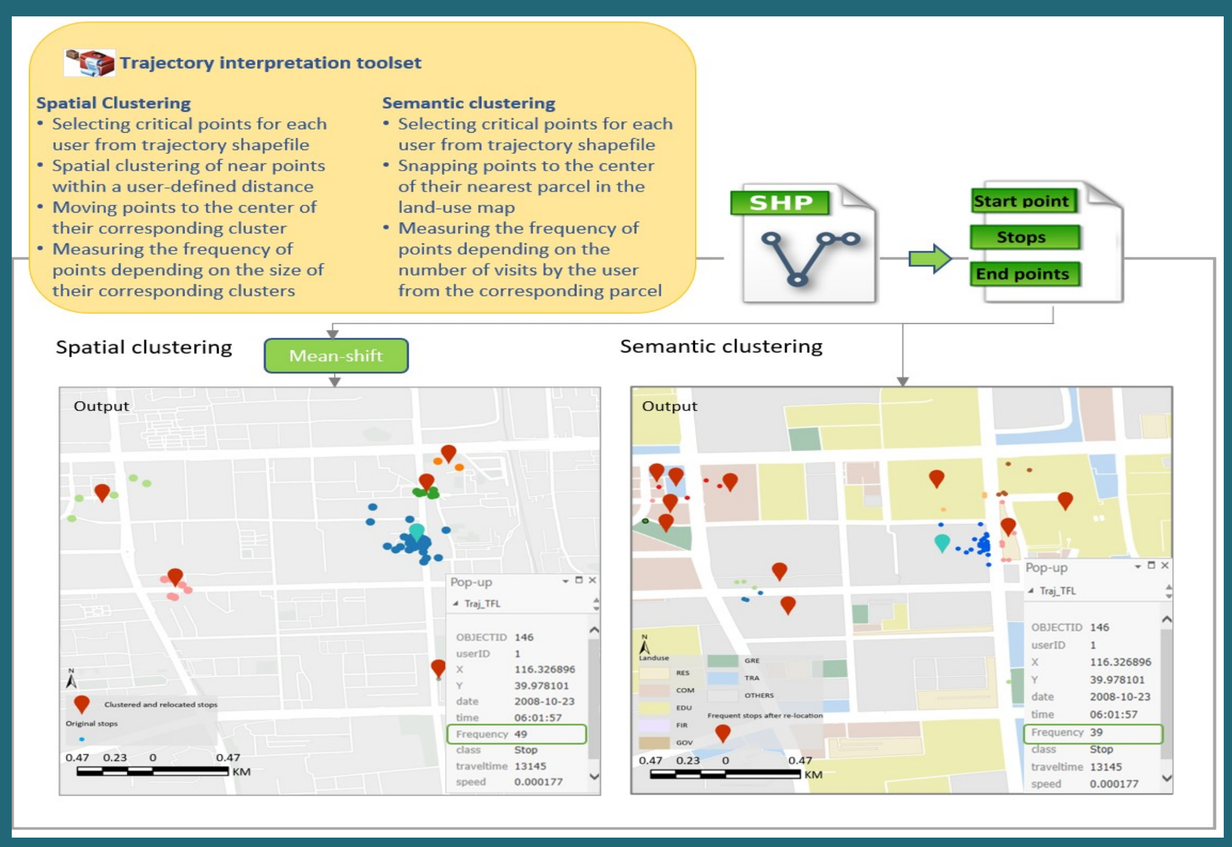
RESULTS





DATASET

To empirically test the performance of the toolbox, a real-world dataset including pseudo-anonymized GPS records from the Geolife project [2], [79], [80] was used. In Geolife, 182 users tracked their daily outdoor movements from 2007 to 2012, mainly in Beijing. The dataset contains timestamped latitude and longitude of 17,621 daily trajectories with a total distance of about 1.2 million kilometers and a total duration of 48,000+ hours. 91.5 percent of the trajectories have a high sampling interval (1~5 seconds). The City of Beijing, China was chosen as the study area since most GPS records are placed in this city.



CONCLUSION

PrivacyProtection toolbox is publicly available as a geoprocessing package on ArcGIS online. However, it must be noted that data anonymization is a difficult that requires significant domain expertise, particularly with trajectory data. Easy-to-use toolsets are essential, but not a silver bullet as they do not replace expert decision-making when selecting anonymization parameters. Some may argue that such frictionless tools may lull ill -equipped users into a false sense of security, and it is for this reason that we have included a warning in the tool's interface about the importance of parameter selection when anonymizing spatial data. Nevertheless, while our tool cannot replace conceptual expertise, it significantly lowers the barrier to actually operationalizing anonymization methods and thus brings them within reach of a wider audience.