Journal of Transport & Health xxx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Journal of Transport & Health

journal homepage: www.elsevier.com/locate/jth

Using crowdsourced data to monitor change in spatial patterns of bicycle ridership

Darren Boss[a,*], Trisalyn Nelson[b], Meghan Winters[c], Colin J. Ferster[d]

[a] University of Victoria, 3800 Finnerty Road, Victoria, BC, Canada V8P 5C2
[b] Arizona State University, 975 S. Myrtle Avenue, Tempe, AZ 85287, United States
[c] Faculty of Health Sciences, Simon Fraser University, Burnaby, BC, Canada V5A 1S6
[d] Department of Geography, University of Victoria, PO Box 1700 STN CSC, Victoria, BC, Canada V8W 2Y2

ARTICLE INFO

Keywords:
Cycling
Crowdsource
Spatial analysis
Networks
Infrastructure

ABSTRACT

Cycling is a sustainable mode of transportation with numerous health, environmental and social benefits. Investments in cycling specific infrastructure are being made with the goal of increasing ridership and population health benefits. New infrastructure has the potential to impact the upgraded corridor as well as nearby street segments and cycling patterns across the city. Evaluation of the impact of new infrastructure is often limited to manual or automated counts of cyclists before and after construction, or to aggregate statistics for a large region. Due to methodological limitations and a lack of data, few spatially explicit approaches have been applied to evaluate how patterns of ridership change following investment in cycling infrastructure. Our goal is to demonstrate spatial analysis methods that can be applied to emerging sources of crowdsourced cycling data to monitor changes in the spatial-temporal distribution of cyclists across a city. Specifically, we use crowdsourced ridership data from Strava to examine changes in the spatial-temporal distribution of cyclists in Ottawa-Gatineau, Canada, using local indicators of spatial autocorrelation. Strava samples of bicyclists were correlated with automated counts at 11 locations and correlations ranged for 0.76 to 0.96. Using a local indicator of spatial autocorrelation, implemented on a network, we applied a threshold of change to separate noise from patterns of change that are unexpected given a null hypothesis that processes are random. Our results indicate that the installation or temporary closure of cycling infrastructure can be detected in patterns of Strava sample bicyclists and changes in one location impact flow and relative volume of cyclists at multiple locations in the city. City planners, public health professionals, and researchers can use spatial patterns of Strava sampled bicyclists to monitor city-wide changes in ridership patterns following investment in cycling infrastructure or other transportation network change.

1. Introduction

Cycling is a sustainable mode of transportation with numerous health, environmental and social benefits (Gordon-Larsen et al., 2005; Pucher and Buehler, 2008; Teschke et al., 2012). In an effort to increase ridership, many cities are making significant financial investments in cycling infrastructure, and several cities are developing cycling infrastructure networks (Buehler and Dill, 2015). It is essential that cities monitor and report on the impact of infrastructure projects on ridership to be accountable to the public and to

* Corresponding author. Permanent address: PO Box 948, Cumberland, BC, Canada V0R1S0.
E-mail addresses: dboss@uvic.ca (D. Boss), trisalyn.nelson@asu.edu (T. Nelson), mwinters@sfu.ca (M. Winters).

https://doi.org/10.1016/j.jth.2018.02.008
Received 2 November 2017; Received in revised form 10 February 2018; Accepted 16 February 2018
2214-1405/

encourage political will for future investments in cycling infrastructure (Handy et al., 2014).

Monitoring and evaluation of the impacts of investment in cycling infrastructure across a city have been difficult due to a lack of spatially explicit ridership data. Traditional sources of cycling data and route information include manual counts, intercept surveys, automated pneumatic tube counters, mail-back surveys, travel surveys, and research projects (Forsyth et al., 2010; Hyde-Wright et al., 2014; Nordback et al., 2013). While traditional cycling data provide important information on bicycling levels, standard data lack the spatial and temporal detailed needed for mapping change in bicycling levels. Data limitations are being overcome through advances in Global Positioning System (GPS) technology and its incorporation into portable devices, such as smartphones, which provides a novel source of spatially and temporally dense cycling data. Further, large citizen science cycling datasets are becoming available for analysis (Romanillos et al., 2015). For example, in North American cities, crowdsourced cycling data has been used to examine where cycling for health occurs with respect to land use diversity, bicycle facilities and residential and employment density (Griffin and Jiao, 2015) and researchers have observed a strong correlation between crowdsourced fitness app data and manual cycling counts (Jestico et al., 2016).

In order to utilize crowdsource data for monitoring ridership change we must identify suitable analytical methods. Spatial statistics enable spatial patterns in data to be mapped and unusual patterns identified (Nelson and Boots 2008). When mapping change, it is essential to differentiate minor and random change from substantive change. A group of spatial statistics, local measures of spatial autocorrelation, can be used to quantify relatedness in nearby events and to map when and where patterns are statistically associated with non-random spatial processes (Anselin 1995). Specifically Local Moran's I_I is used to map clusters of extreme change.

Our goal is to demonstrate how spatial pattern methods can be applied to crowdsourced ridership data to monitor changes in the spatial-temporal variation of ridership across a city. We analyzed a large crowdsourced cycling dataset for Ottawa-Gatineau, Canada, comparing volumes of cyclists from May 2015 and May 2016 to meet the following objectives. First, we evaluate the appropriateness of using crowdsource data to represent bicycling levels. Second, we quantified change in patterns of ridership using network appropriate measures of local spatial autocorrelation. Third, we tracked changes associated with three bicycling infrastructure projects that occurred over the time period of our study.

2. Study area and data

2.1. Study area

The case study area is Ottawa-Gatineau, Canada with a population of 1.24 million (Statistics Canada, 2011a). Approximately 2.2% of workers commute by bicycle (Statistics Canada, 2011b). From 2006 to 2011 daily bicycle trips grew from 30,350 to 43,350 (an increase of 43%) (City of Ottawa 2013). The region has invested significant financial resources in bicycle and multi-use infrastructure over the past several years and currently has over 600 km of bicycle paths (National Capital Commission, 2017). Infrastructure that we monitor for change in ridership patterns: are Adawe bike and pedestrian bridge (opened December 2015), Hickory bike and pedestrian bridge (opened August 2015), MacDonald-Cartier pathway (opened December 2015).

2.2. Official Bicycling Data

Ottawa-Gatineau had 10 bicycling counters in 2015 and 11 in 2016 and we utilized bicycle counts on weekdays (24-h days) in May for 2015 and 2016. Bicycling counters are automated counters located throughout the city. Data are reported daily and counters are considered accurate within +0 and −5% of bikes that cross the sensing section of the pathway. Data are accessed from an open data portal managed by the City of Ottawa. We compare official data with crowdsourced data described below.

2.3. Crowdsourced Bicycling Data

The City of Ottawa has partnered with Strava, a social network for runners and cyclists, to obtain a large crowdsourced cycling dataset. The Strava mobile App is used by athletes to track their activities which are then uploaded to the Strava website. This volunteer sourced data is anonymized and aggregated into the Strava Metro data product (Strava Metro, 2017). The Strava Metro data used in this study consists of activity counts (bicycle trips) per segment of transportation infrastructure in the Ottawa-Gatineau region, aggregated by month using weekday data in May 2015 and May 2016. We chose this time period as several substantial changes were made to cycling infrastructure between May 2015 and 2016, and thus this serves as a pre-post analysis. There were a total of 4.49 million activity counts from 52,123 bike trips across 71,205 network segments. Strava is used most commonly by recreational cyclists but in dense urban areas correlates with all bicyclists (Jestico et al. 2016). In Ottawa-Gatineau we expect a higher proportion of commuters than typical in Strava data due to a marketing campaign led by the city for commuters to contribute data to Strava in advance of the data purchase. The street segment map included in the Strava Metro data product was derived from OpenStreetMap (OpenStreetMap, 2017).

The demographics of the Strava users in Ottawa-Gatineau are not representative of the general cycling population, there are differences in both gender and age. The percentage of male Strava users (78.2%) is higher than the percentage of male cyclists in the Ottawa-Gatineau region (68%) (TRANS Committee, 2011). Strava users in the 25–34 and 35–44 age groupings are over-represented as compared to the actual cycling population, while the under 25, 55–64, 65–74 and over 75 age groupings of Strava users are under-represented (Fig. 1). The trends in the Strava data used in this study are very similar to age and gender trends of crowdsourced data used in other bicycling studies (i.e., Griffin and Jiao 2015; Romanillos et al. 2016).