

Photos, tweets, and trails: Are social media proxies for urban trail use?

Xinyi Wu

University of Minnesota
wuxx1088@umn.edu

Greg Lindsey

University of Minnesota
linds301@umn.edu

David Fisher

Stanford University and University of
Washington
davefisher@stanford.edu

Spencer A. Wood

Stanford University and University of
Washington
woodsp@stanford.edu

Abstract: Decision makers need information on the use of, and demand for, public recreation and transportation facilities. Innovations in monitoring technologies and diffusion of social media enable new approaches to estimation of demand. We assess the feasibility of using geo-tagged photographs uploaded to the image-sharing website Flickr and tweets from Twitter as proxy measures for urban trail use. We summarize geo-tagged Flickr uploads and tweets along 80 one-mile segments of the multiuse trail network in Minneapolis, Minnesota, and correlate results with previously published estimates of annual average daily trail traffic derived from infrared trail monitors. Although heat maps of Flickr images and tweets show some similarities with maps of variation in trail traffic, the correlation between photographs and trail traffic is moderately weak (0.43), and there is no meaningful statistical correlation between tweets and trail traffic. Use of a simple log-log bivariate regression to estimate trail traffic from photographs results in relatively high error. The predictor variables included in published demand models for the same trails explain roughly the same amount of variation in photo-derived use, but some of the neighborhood socio-demographic and built-environment independent variables have different effects. Taken together, these findings show that both Flickr images and tweets have limitations as proxies for demand for urban trails, and that neither can be used to develop valid, reliable estimates of trail use. These results differ from previously published results that indicate social media may be useful in assessing relative demand for recreational destinations. This difference may be because urban trails are used for multiple purposes, including routine commuting and shopping, and that trail users are less inclined to use social media on trips for these purposes.

Keywords: trail use, social media, Flickr, Twitter, geo-tagging, recreation, transportation

Article history:

Received: November 26, 2016

Received in revised form: April
17, 2017

Accepted: June 29, 2017

Available online: October 10,
2017

Copyright 2017 Xinyi Wu, Spencer A. Wood, David Fisher & Greg Lindsey

<http://dx.doi.org/10.5198/jtl.u.2017.1130>

ISSN: 1938-7849 | Licensed under the [Creative Commons Attribution – Noncommercial License 3.0](https://creativecommons.org/licenses/by-nc/3.0/)

The *Journal of Transport and Land Use* is the official journal of the World Society for Transport and Land Use (WSTLUR) and is published and sponsored by the University of Minnesota Center for Transportation Studies.

1 Introduction

Policymakers, facility managers, and program evaluators need information about the demand for, and use of, public facilities. These data are essential for purposes ranging from prioritizing investments to assessing safety risks to planning maintenance schedules. Obtaining information about use of public facilities, particularly transportation and recreation facilities that are open-access, is challenging, historically requiring long hours of observation or surveys of users that are difficult to administer (Cranshaw, Schwartz, Hong, & Sadeh, 2012). With the development of new technologies, however, analysts are assessing and implementing new methods of measuring demand. In the transportation and recreation fields, these methods range from deployment of specialized automated traffic monitors (Federal Highway Administration, 2013; Minge, Falero, Lindsey, & Petesch, 2015) to acquisition and interpretation of measures of social media use that can be interpreted as proxies for demand (Wood, Guerry, Silver, & Lacayo, 2013). Choices among these new approaches involve tradeoffs between capital, operating, and maintenance costs and the quality of data obtained for planning and decision making. Big data from social media are particularly attractive because they can be obtained and analyzed without the labor and expense of deploying new capital equipment specifically for monitoring. Researchers have shown that social media can be used to assess demand in some recreational destinations, but additional studies are warranted because of the many different contexts in which estimates of demand are needed.

This paper explores the feasibility of using measures of social media as proxies for demand for urban multiuse trails. Urban trails, or shared use paths, which often are built along rivers or historic rail lines, form the backbone of non-motorized transportation networks in metropolitan areas, connecting residential areas to both recreation destinations such as parks and lakes and economic and employment centers. Demand for these facilities is growing; the Rails to Trails Conservancy, a major advocate for trail development, is working with local governments across the United States to develop trail facilities within three miles of 90% of the country's urban population. Development of efficient, unobtrusive methods for assessing demand can inform these types of initiatives.

Researchers have shown that social media can be used to gauge demand at remote parks and other destination recreation facilities (Keeler et al., 2015; Sessions, Wood, Rabotyagov, & Fisher, 2016). Urban trails are interesting facilities and present additional challenges for monitoring because they are open-access, have multiple formal and informal access points, are integrated with other transportation and recreation infrastructure, and are used for both recreation and utilitarian purposes such as commuting or shopping. Because trail use for these utilitarian purposes is more likely to be routine and less likely to involve novel experiences that users may want to share with friends, use of social media may be less likely than at purely recreational facilities. It follows that measures of social media use may not correlate as well with overall facility use.

We begin our study with a brief discussion of relevant background research, focusing on recent advances in methods for estimating demand for transportation and recreation facilities (Section 2). Next, we describe our study area, data, and methods (Section 3). We present descriptive and analytic results using automated, infrared counts of trail traffic to assess and validate measures of demand derived from social media, specifically geo-tagged photographs uploaded to Flickr and tweets shared through Twitter (Section 4). We conclude that these social media cannot be interpreted as valid proxies of demand for urban multiuse trails (Section 5).

2 Background and literature review

Policymakers and program managers have long needed measures of demand to plan, build, operate, and maintain public facilities. Traffic monitoring results are used for multiple purposes, ranging from

prioritizing new facilities to allocation of federal funding for maintenance of existing facilities. In the field of recreation, public agencies collect data on facility use for the same types of reasons. The Metropolitan Council in the Twin Cities Metropolitan Area in Minnesota, for example, annually observes and counts trail visitors at dozens of trail access points throughout the seven local counties and uses results to allocate funding (Metropolitan Council, 2016). With technological advances, program managers are exploring new ways to monitor use and gauge demand for facilities.

One approach to assessing demand involves deployment of new specialized automated monitoring devices designed to collect data at particular types of facilities. The FHWA (Federal Highway Administration, 2013) recently published new guidance on the use of infrared, radio beam, inductive loop, pneumatic tube, and other devices developed specifically to count bicyclists and pedestrians on public infrastructure. States and many local jurisdictions now are implementing programs, although these remain in their infancy (Minge et al., 2015; Griffin, Nordback, Gotschi, Stoltz, & Kothuri, 2014; Jackson, Stolz, & Cunningham, 2015; Lindsey, Petesch, & Hankey, 2015). These efforts typically involve establishment of a set of permanent automated counters at a fixed number of locations believed to represent specific traffic patterns and then collection of a large number of short-duration (e.g., two to seven day) samples that can be used to characterize flows on a network. Statistics and ratios from the permanent sites are used to extrapolate the sample counts and obtain estimates of daily traffic (Federal Highway Administration, 2013; El Esawey, Lim, Sayed, & Mosa, 2013; Miranda-Moreno, Nosal, Schneider, & Proulx, 2013; Nordback, Marshall, Janson, & Stolz, 2013; Hankey, Lindsey, & Marshall, 2014). Several studies of trail use using these technologies and monitoring approaches have been reported (e.g., Lindsey, Wilson, Rubchinskaya, Yang, & Han, 2007; Wang, Lindsey, Hankey, & Hoff, 2014). Wang, Hankey, Wu, and Lindsey (2016) illustrated how these procedures can be used to characterize and model non-motorized traffic flows on urban multiuse trail networks.

A second approach to assessing demand involves strategies to capture information about facility use from surrogate sources not designed specifically for the purposes of monitoring facility use. Perhaps the best example of this approach is the growing use of social media as a proxy of demand for particular activities or facilities. The approach has become feasible because the use of these media has penetrated deeply into society and become mainstream. Simultaneously, their use for analysis has grown, and the types of applications have diversified. Some researchers are developing new algorithms to exploit geo-referenced data in social media, while others are focusing on substantive applications, including estimation of facility use. Backstrom, Sun, and Marlow (2010) developed an algorithm to identify the location of individuals based on Facebook networks among friends. Lenormand, Tugores, Colet, and Ramasco (2014) analyzed tweets along highway and railway segments in 39 European countries to illustrate variations in social media use in association with these transportation modes. Some researchers have explored social media data in tourism and hospitality services industries (Miguens, Baggio, & Costa, 2008; Verma, Stock, & McCarthy, 2012).

In the field of transportation, researchers are exploring the use of data from fitness applications and other custom-designed applications for smart phones to track cyclists and analyze their behaviors, including route choice. Griffin and Jiao (2015) used Strava data to identify routes chosen by cyclists for fitness (i.e., health) rides. Jestico, Nelson, and Winters (2016) found that crowdsourced counts of cyclists from Strava had moderate correlation with manual counts (R^2 0.40 to 0.58), with one crowdsourced cyclist representing 51 observed cyclists. They noted the bias in these data: fitness applications like Strava are used mainly by elites. They concluded, however, the data could be used to predict categories of ridership and illustrate spatial variation in route selection.

As for recreation, Wood et al. (2013) conducted a study using photos on Flickr to estimate the visitation rates at recreational sites around the world. This study demonstrated the feasibility of using

social media as a proxy for visitation at sites where measures of demand previously were very difficult to obtain because of their locations and characteristics. In a regional assessment based upon these methods and findings, Keeler et al. (2015) used photographs from Flickr to estimate demand for recreational use of lakes in Minnesota. They also estimated regression models to identify factors affecting visitation rates to the lakes in their sample. These studies confirmed the feasibility of using social media data to estimate visitation rates at destinations. Researchers also have begun to use social media to estimate demand for differentiated services or features offered at large facilities. For example, Richards and Friess (2015) classified the photos on Flickr based on their contents, allowing researchers to differentiate and estimate the use of different cultural services such as nature appreciation, social recreation, and fishing.

One reason for the success of these efforts is that destination recreation sites share characteristics that make use of social media likely and amenable to analysis. By definition, most visitors to destination recreation facilities visit voluntarily for purposes of leisure and have time and reasons to share their experience via social media. In comparison, the use of social media during use of public facilities in urban areas that serve multiple purposes, including those with characteristics that mediate against use of social media, is not as well understood. Studies of the potential to use social media as proxies of demand for use of multi-purpose urban public facilities therefore are warranted.

3 Study area, data, and methods

Our principal objective is to assess whether geo-referenced data from social media, specifically Flickr and Twitter, can be used as proxies for demand for urban trail facilities. We choose photographs uploaded to Flickr and tweets because both measures have been shown to be correlated with use of recreational facilities and because many, if not most, people use urban multiuse trails for recreational purposes. Our study area is the City of Minneapolis, Minnesota, where comprehensive automated trail traffic monitoring using infrared devices (Lindsey et al., 2015; Wang et al., 2016) enables comparison and validation of measures of demand derived from social media.

Following methods developed by Wood et al. (2013), we extract and analyze tweets and Flickr posts in 80 one-mile segments of trail corridors within the city for a one-year period. Using log-log correlation models, we correlate these measures with estimated annual average daily trail traffic (AADTT) reported by Wang et al. (2016). Next, adapting multivariate trail demand models developed by Wang et al. (2016), we test how well the models explain variation in tweets and Flickr posts along trail segments. These models include neighborhood socio-demographic, built environment, and trail location environment variables.

3.1 Social media: Counts of Flickr uploads and tweets

Following methods pioneered by Wood et al. (2013) and used by Keeler et al. (2015), we obtained counts of Flickr images and tweets from locations within buffer zones along the 80 miles of trails in Minneapolis for the calendar year 2013, the year for which complete trail monitoring results are available. This process involved buffering the trail network using standard GIS procedures and then extracting geo-tagged Flickr images and tweets for each trail segment. We used a 200-foot buffer for the trail segments for both forms of social media. To test sensitivity to scale, we also experimented with a 50-foot buffer. After buffering the trail segments, we use the polygons to spatially query the sets of geotagged photos and tweets. For Flickr photos, we access the global dataset via the Flickr API, and conduct the spatial queries using the InVEST Visitation model (Sharp et al., 2016). For tweets, we access the data via the Twitter Public Stream API. Latitude/longitude coordinates associated with photos and tweets are typically generated by the camera or mobile device that created the content, and are accurate to within

tens of meters.

We report results for Twitter analyses; the results for the Flickr uploads from the 50-foot buffer are not reported because the numbers of uploads were too low for analysis. Details on the extraction procedures are reported in Wood et al. (2013).

The measures we test as proxies for trail demand are photo-user days (PUDs) and twitter-user-days (TUDs). The total PUDs for a single segment is the number of unique combinations of user and date (of photo taken) from the set of photos returned in the spatial query for that segment. The calculation of TUDs is identical with the calculation of PUDs. For example, if an individual along a trail segment uploaded multiple photos or tweets from the same trail segment on the same day, this would be reported as a single PUD or TUD. We averaged the PUDs and TUDs for each trail segment for each day in 2013. Our two resulting measures, annual average photo-user days (AAPUD) and annual average Twitter-user days (AATUD) are conceptually analogous to our measure of trail use, annual average daily trail traffic (AADTT).

3.2 Trail traffic counts

We use estimates of trail traffic previously reported by Lindsey et al. (2015) and Wang et al. (2016) as the validation measure to assess whether social media posts are correlated with trail traffic. The trail traffic monitoring program in Minneapolis was initiated by the Minneapolis Park and Recreation Board, the Minneapolis Department of Public Works, and the University of Minnesota following guidelines established by the Federal Highway Administration (2013). Estimates of trail AADT were made using field collection and extrapolation procedures developed by Hankey et al. (2014) and described in Lindsey et al. (2015). Their monitoring network included 80 trail segments that averaged about one mile in length. The monitoring network included six reference monitoring locations where counts were taken for the entire year. All short-duration counts were taken for a minimum of seven complete days between April and October. Mixed mode (undifferentiated bicyclist and pedestrian) trail traffic counts were made using permanent and portable active infrared sensors. Active infrared counters systematically undercount trail traffic because of occlusion (i.e., multiple users passing the sensor simultaneously). To assess the magnitude of undercount, researchers conducted 130 hours of manual validation counts, computed a correction equation, and adjusted all hourly counts to correct for the systematic undercount (Hankey et al., 2014). The correction equation was: $y = -0.0002x^2 + 1.3695x - 5.8067$ (1) (Adj $R^2 = 0.9923$). All estimates of annual average daily trail traffic were computed using the corrected counts. The day-of-year factoring method was used to extrapolate sample counts (Hankey et al., 2014; Lindsey et al., 2015). This method is based on the assumption that traffic patterns at the sample sites are similar to those at the reference sites, specifically, that the proportion of annual traffic accounted for by the sample count on each segment is the same as the proportion of annual traffic accounted for by those same days at the reference sites.

3.3 Analysis and modeling

We summarize descriptive statistics to illustrate relative volumes of Flickr images, tweets, and trail traffic and present heat maps to illustrate spatial variation in annual average PUDs, TUDs, and daily trail traffic along trail segments. We use simple log-log, bivariate regression to test correlation between the geo-tagged social media measures and trail traffic. To assess whether photo-user days can be used to estimate traffic, we estimate AADTT for each trail segment from AAPUD regression equations, and then calculate the mean absolute percentage error (MAPE) for the estimates.

We also present a set of three negative binomial regression models to illustrate factors associated

with AADTT and assess whether these same factors are correlated with photo-user days and Twitter-user days. Negative binomial regression is especially useful when the dependent variable is a non-negative integer such as a count, and these models have been used widely to model traffic counts and volumes (e.g., Sessions et al., 2016). Our base model is the model of AADTT previously presented by Wang et al. (2016). Their model controls for neighborhood socio-demographics, the built environment, and relevant trail characteristics. Descriptive statistics for explanatory variables are presented in Table 1. We use the Cox-Snell R-squared as an indicator of the goodness of the fit of the models. All models were estimated in STATA.

Table 1: Variables included in trail and social media demand models (Wang et al., 2016)

Variables	Description	Mean (Std. Dev.)	Expected Effect on Non-motorized Traffic
Socioeconomic			
dPct_yo	% people with age under 6 or over 65	0.16(0.08)	Positive/Negative
Pct_Others	% other ethnicity, exclude White and African American	0.13(0.11)	Negative
Pct_black	% African Americans	0.1(0.15)	Negative
M-income	Median Household Income	68000(31520)	Positive
Built Environment			
PopDen	Gross population density (people/acre)	8.41(6.88)	Positive
Diversity	5-tier employment entropy	0.5(0.3)	Positive
NetDen_Ped	Network density in terms of facility miles of pedestrian links per square mile	19.8(6.62)	Positive
IntDen_Auto	Intersection density in terms of auto-oriented intersections per square mile	4.16(9.08)	Positive
Pct_ind	% Industry	0.16(0.22)	Negative
Job_byW	Number of jobs accessed by 20-min walk	9480(20420)	Positive
Trail Location Characteristics (Dummy)			
Disconnect	Disconnected segment or not	0.12(0.32)	Negative
Lake	In the lake area or not	0.19(0.40)	Positive

4 Results

4.1 Descriptive results

Estimates of annual average daily trail traffic, photo-user days, and Twitter-user days across the 80 trail segments are summarized in Table 2. Because not all trail users take and upload photos to Flickr or send tweets, it is expected that AADTT will be higher than both AAPUD and AATUD. This outcome is clear in the data. Across the 80 trail segments, annual average daily trail traffic ranges over three magnitudes, from a low of 39 to a high of more than 3,700, with a mean of approximately 950. AAPUD is much lower, ranging from a minimum of less than 1 to 166 with a mean of about 18 (i.e., average daily trail traffic is more than 50 times higher). Compared to Flickr uploads, there were many more tweets. For the 200-foot buffer, AATUD ranged from 11 to more than 2,300, with a mean of about 310. The mean number of Twitter-user days per segment for the 200-foot buffer was approximately four times the mean for the 50-foot buffer. The coefficient of variation for the photo-user days (1.65) was higher than for the other three measures, indicating greater variation or dispersion of uploads across the network.

Table 2: Trail AADTT, AAPUD, and AATUD in Minneapolis

Variable	Observations	Mean	Std. Dev.	Min	Max	Median
Annual Average Daily Trail Traffic (AADTT)	80	953.8	971.9	39.5	3727.9	750.2
Annual Average Photo-User-Days (AAPUD) (200-ft buffer)	80	17.6	29.0	0.5	165.8	7.9
Annual Average Twitter-User-Days (AATUD) (200-ft buffer)	80	309.81	434.25	11	2309.5	148.5

To illustrate spatial variation in use of social media we created heat maps (Figure 1b, c, d). These maps bear some similarities to the map of trail traffic (Figure 1a). Both trail traffic and use of social media are higher around the series of lakes and the multi-use trails that connect them in south and southwest Minneapolis, although there is variation among the lakes. In particular, the use of social media is lower around Lake Nokomis, the southern-most lake, than around the other lakes to the west and north, each of which is in higher-income neighborhoods. Similarly, trail traffic and both social media measures are lower in north Minneapolis where there are fewer lakes that serve as recreation destinations, some trails are not connected to the network, and neighborhoods generally are poorer. Spatial patterns diverge in some areas, particularly in south Minneapolis where trail traffic on a trail adjacent to a light rail line and an arterial highway is relatively low but twitter use is relatively high. This result may reflect the presence of the light rail train. Finally, the use of social media is high relative to trail traffic around the campus area of University of Minnesota. The areas in north Minneapolis where trails are not very well connected have less trail traffic and relatively fewer social media posts. North Minneapolis is also one of the poorest areas of the city.

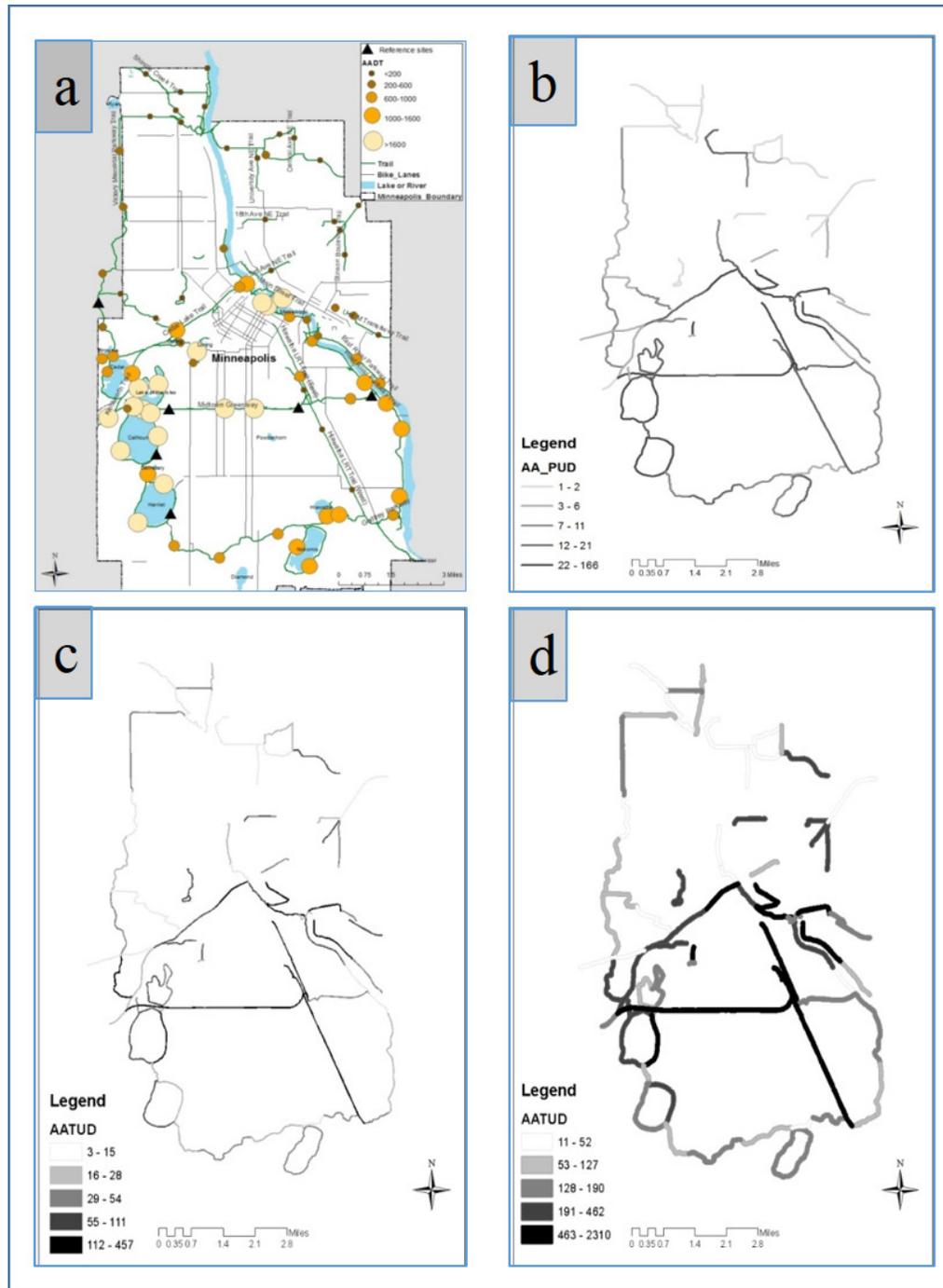


Figure 1: Maps of annual average daily trail traffic (AADTT), annual average photo-user days (AAPUD), and annual average Twitter-user days (AATUD)

4.2 Analytic and regression results

This study uses log-log linear models between measures of trail traffic and social media use to examine their degree of correlations. Regression diagnostics indicated the log-log approach controls for non-linearities meets the distributional assumptions of OLS regression. When plotted against independent variables, residuals are relatively symmetrically distributed around zero, without showing any clear patterns.

The regression results show there is positive but moderately-weak correlation between photo uploads and trail traffic but no meaningful correlation between tweets and trail traffic. Specifically, the annual average photo user-days explained approximately 43% of trail AADT ($R^2 = 0.43$) but Twitter-user days explained only 5% and 7% at the 200 and 50 meter buffers, respectively. (Table 3; Figure 2). While t-tests for each of the social media variables are highly significant, and F-tests indicate good fit for each model overall, the results indicate that these measures have limited value as proxies for trail traffic.

Table 3: Hypothesis testing of the coefficients and the models

Dependent variable	Independent variable	t-test (P-value)	Prob - F
Log AADTT (Annual Average Daily Trail Traffic)	Log PUD (Photo-user-days)	0.000	0.000
	Log TUD (Twitter-user days) (50 ft)	0.051	0.0505
	Log TUD (Twitter-user days) (200 ft)	0.022	0.0216

To assess further whether AAPUD (which had the best fit) might serve as a proxy for trail traffic, we used the log-log regression equation to predict AADTT and computed the mean absolute percentage error (MAPE) for the 80 estimates. The MAPE was 124%. The error was less than 25% for 15% of the estimates; it was more than 100% for 20% of the estimates (Figure 3). This magnitude of error is additional evidence that, in this trail network, photo-uploads to Flickr cannot be used to predict trail traffic accurately. We did not estimate trail traffic from tweets because correlation between Twitter posts and trail traffic was so low.

Wang et al. (2016) showed using negative binomial regression that much of the variation in traffic across trail segments can be explained by neighborhood socio-demographics, the built environment, and trail characteristics contiguous to individual trail segments (Cox-Snell $R^2 = 0.64$; Table 4). As a way of confirming the results of the log-log regressions, we explored the extent to which the variables in the Wang et al. (2016) trail model explain variation in photo uploads and Twitter use (Table 4). That is, our goal was to see if the variables that explain trail use also explain use of social media, not to develop the best models of social media use. Consistent with other analyses, the variables in the trail demand model explain more of the variation in photo uploads variation in AAPUD (Cox-Snell $R^2 = 0.67$) than use of Twitter, regardless of buffer width (Cox-Snell R^2 s = 0.33/0.39).

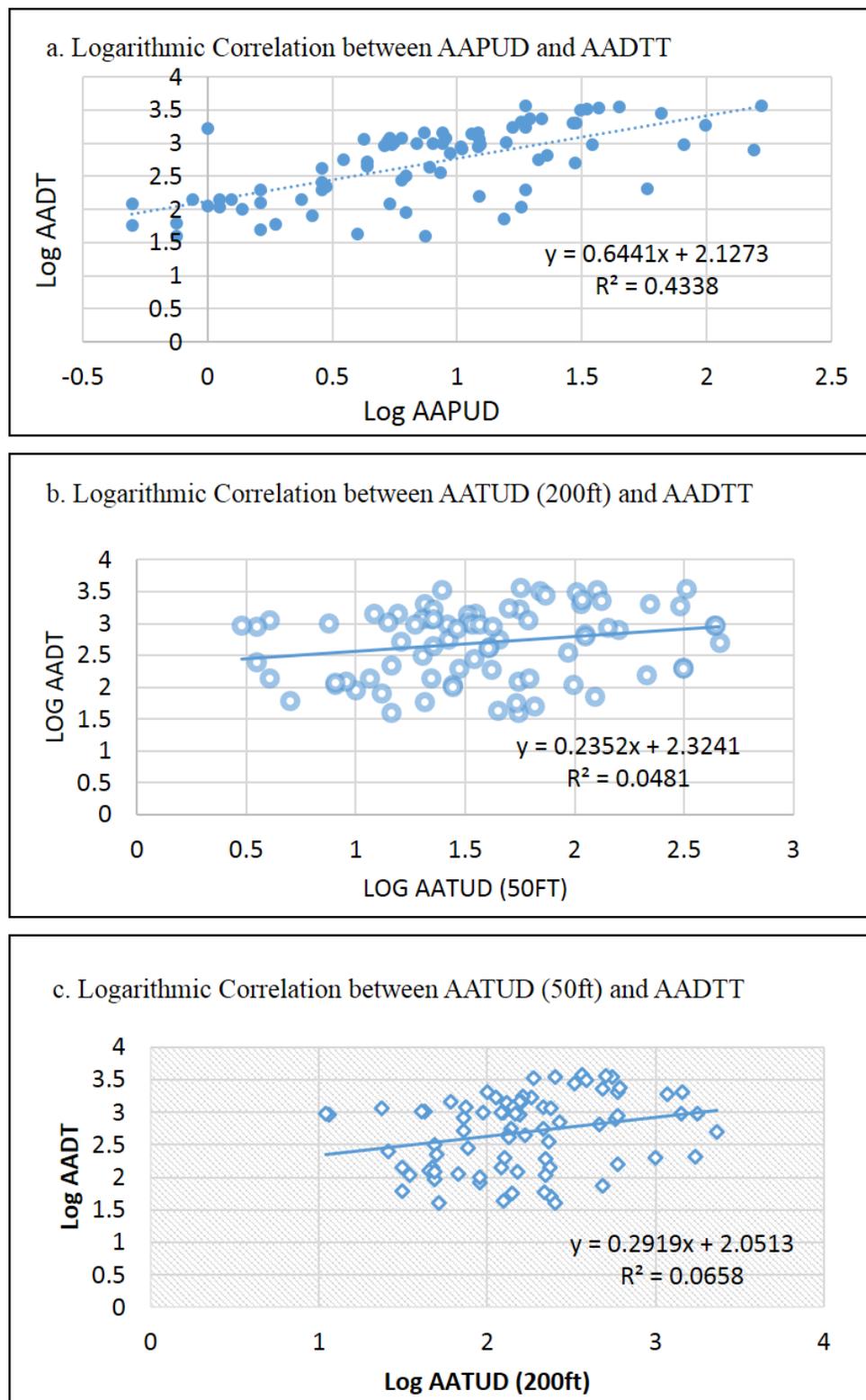


Figure 2: Log-log regression of annual average photo-user days (AAPUD), annual average Twitter-user days (AATUD) and annual average daily trail traffic (AADTT)

Table 4: Modeling results

Variables	Annual Average Daily Trail Traffic (AADTT) (16)		Annual Average Photo-User-Days (AAPUD)		Annual Average Twitter-User-Days (AATUD) (50-ft buffer)		Annual Average Twitter-User-Days (AATUD) (200-ft buffer)	
	Coefficient	P	Coefficient	P	Coefficient	P	Coefficient	P
Constant	5.51	0.00	2.03	0.663	4.48	0.000	5.58	0.000
Socioeconomic								
% of people with age under 6 or over 65	0.48	0.674	0.62	0.678	-0.74	0.672	-1.40	0.388
% of other ethnicity, exclude White and African American	-1.49	0.028	-1.03	0.967	-0.76	0.477	-0.27	0.792
% of African Americans	-2.72	0.003	-1.39	0.099	-1.24	0.254	-1.65	0.085
Median Household Income	0.0005	0.882	0.00001	0.003	-9.82e-06	0.058	-9.69e-06	0.040
Built Environment								
Gross population density (people/acre)	0.05	0.003	0.003	0.883	0.05	0.080	0.04	0.097
5-tier employment entropy	1.05	0.001	1.19	0.001	0.50	0.239	1.001	0.011
Network density in terms of facility miles of pedestrian links per square mile	0.02	0.133	0.03	0.157	-0.004	0.879	0.011	0.613
Intersection density in terms of auto-oriented intersections per square mile	0.00	0.773	0.01	0.371	0.004	0.799	0.005	0.747
% of Industry	-0.36	0.477	-1.09	0.080	-0.22	0.774	-0.56	0.450
Number of jobs access by 20 min walk	0.004	0.378	0.00003	0.000	0.00001	0.083	7.24e-06	0.349
Trail Location Characteristics (Dummy)								
Disconnected segment or not	-1.39	0.000	-1.33	0.000	-0.46	0.418	-0.249	0.538
In the lake area or not	1.08	0.000	0.80	0.001	0.13	0.686	-0.10	0.733
Model Fit								
Cox_Snell R ²	0.64		0.67		0.33		0.39	
P in LR test	0.000		0.000		0.000		0.000	

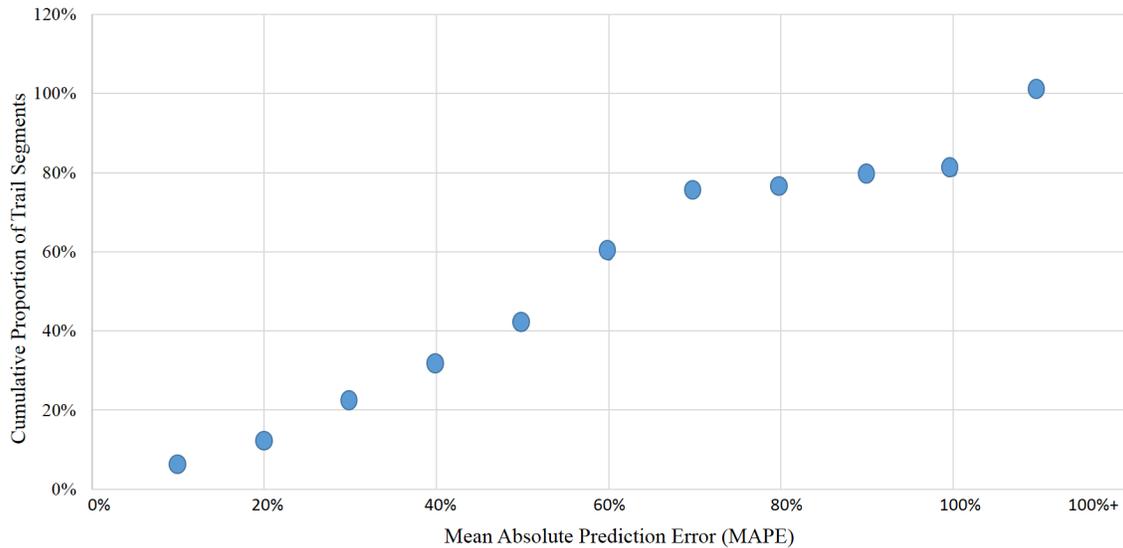


Figure 3: Error in estimating annual average daily trail traffic (AADTT) from annual average photo-user days

The social media models have comparable fit, but different variables are significant in each equation. Among the four neighborhood socio-demographic variables, the biggest difference is associated with the income variable, which is not correlated significantly with Twitter use but is significantly correlated with photo uploads (P -value = 0.003). Use of both types of social media are lower in neighborhoods with higher populations of black residents, although the effect of this variable is barely significant at the 10% level in the AAPUD model (P -value = 0.099). The proportion of residents between 5 and 65 in contiguous neighborhoods has no effects. Among the neighborhood built environment variables, employment diversity has the highest significant, positive effect on both measures of social media (P -value = 0.001 in each model). The effects of other variables diverge. Population density is highly correlated with trail traffic but not photo uploads while the number of jobs within a 20-minute walking distance is significantly correlated with photo uploads but not trail traffic. The network density of pedestrian links and the intersection density are not significant in either model. The two trail location variables were highly significant in both models. Segments contiguous to lakes were correlated positively and significantly with both trail traffic and photo uploads while both trail traffic and Flickr uploads were significantly lower on segments not connected to the network. Despite the differences between the p -values in the trail traffic and photo-user day models, the estimated coefficients mostly have the same sign and are of the same order of magnitude. Overall, however, the divergence among models is evidence that different factors are correlated with trail use and the use of social media. To sum up these regression results: photo uploads and tweets along the Minneapolis trail network have moderate to weak or no correlation with trail traffic; trail traffic estimated from photo-user days results in errors too high for practical applications, and the socio-demographic, built environment, and infrastructure variables that predict trail traffic do not have the same explanatory power for photo-user days or Twitter-user days.

5 Discussion and conclusions

Policy makers and public managers need information about use of public facilities to inform decisions

and are exploring diverse strategies to obtain useful information about demand. These strategies range from establishment of comprehensive monitoring networks of specialized sensors to reliance on surrogate measures from data sources that can be accessed without deployment of new monitoring devices. Decisions to implement programs are affected by the costs and quality of information generated with different approaches. Researchers have shown that geo-tagged social media can be used to assess demand for remote or destination recreation sites. Less is known about the feasibility of using social media as proxies for urban, multipurpose facilities like urban trails.

In this paper, we extracted geo-tagged posts on Flickr and Twitter along mile-long segments of an 80-mile trail network in Minneapolis and assessed how well they correlate with measures of trail traffic obtained from a network of infrared sensors deployed in accordance with guidelines established by the FHWA. Our analyses showed:

- Annual average daily trail traffic on trail segments was, as expected, much higher than annual average Twitter-user days and photo-user days, respectively. This was expected because not all users upload photos or send tweets.
- More people along trails tweet than upload photo images: AATUD along trail segments was higher than AAPUD.
- Mapping of the distribution of trail traffic, photo uploads, and Twitter use along the trail network showed higher values near recreational lakes, lower counts on unconnected trail segments, and other similarities, but also dissimilarities in a major transportation corridor.
- The correlation between AAPUD and AADTT was moderately weak, and use of photo-user days to estimate annual average daily trail traffic was associated with high error.
- Twitter use, regardless of buffer width, was not meaningfully correlated with trail traffic.
- Variables in published trail demand models explain roughly the same amount of variation in photo uploads, but some of the neighborhood control variables have different effects. Neighborhood median household income was not correlated significantly with Twitter use but was significantly and positively correlated with photo uploads. The percentage of black residents in adjacent neighborhoods was associated with lower use of social media.
- Variables in trail demand models have less explanatory power in models of AATUD, and different effects than in AADTT models.
- Our measures of social media performed less well as proxies for facility use than measures for destination recreation sites.

In sum, our analyses indicate that the two forms of social media we tested (i.e., photo uploads and tweets) do not work well as proxies for or predictors of urban trail demand. These findings diverge to some degree from others reported previously in the literature. Compared to estimates of bicycle traffic volume derived from Strava data (Griffin & Jiao, 2015), there is greater error associated with using Flickr photo-user days to estimate trail traffic. This outcome likely occurs because of the relative validity of the two criteria measures. More simply, despite their bias, Strava counts of cyclists are likely to be better correlated with counts of all cyclists than photo uploads are to be correlated with overall trail traffic because the activities are more closely related.

Compared to similar models of lake visitation estimated from Flickr photo-user days (2015), the trail demand models have a lower goodness of fit. Wood et al. (2013) previously speculated that a potential limitation of using social media to explore use of public facilities is that local users tend to post less about places they visit daily. We hypothesize our models have less explanatory power than models of visitor use of recreational destinations because urban trails serve multiple purposes, including routine commuting, shopping or fitness activities, which diminish the novelty of the experience and, with it, the inclination to share.

A potential limitation of our findings concerns the problems of scale and buffering. For example,

one noticeable difference in the spatial distribution of trail traffic and tweets is along the trail in the transportation corridor in south Minneapolis that includes the light rail. It is possible that our measure of Twitter-user days is affected by the proximity of the light rail. We used a buffer of 200 feet to compute photo-user days but were unable to explore the implications of using a narrower buffer because numbers of images were relatively low. Given the width of this buffer, our measure of photo-user days likely includes some users who were not on the trails.

Although our analyses show that the correlation between photo uploads and urban trail use is only moderately-weak, it may be that additional analysis of photo contents would be fruitful. For example, analyses of photo contents to determine if each photo were related to the trail would enable re-calculation of the number of PUDs and re-estimation of correlations. These analyses were beyond the scope of this study. A useful piece of follow-up research would include analyses of the content of the photographs as a way of validating whether the pictures were uploaded by trail users.

These limitations do not mean, however, that these models have no useful practical applications. Although these measures may not be useful for specific activities such as assessing exposure to risk, they may be useful for other, more general planning purposes. Comprehensive trail monitoring programs remain rare, and most large cities do not measure bicycle or pedestrian traffic. In these locations, proxy measures derived from Flickr images may inform efforts to design monitoring programs. These methods also may be extended to rural, destination trails where comprehensive monitoring programs are unlikely to be implemented.

Acknowledgements

We thank the University of Minnesota Institute on the Environment (IonE) for helping to finance this research. Marie Donahue and Bonnie Keeler of IonE provided encouragement and useful insights on these analyses. The Minnesota Department of Transportation, the Minneapolis Park and Recreation Board, and the Minneapolis Department of Public Works collaborated in various elements of this research. This work was supported by the National Socio-Environmental Synthesis Center (SESYNC) under funding received from the National Science Foundation DBI-1052875. None of those individuals or agencies acknowledged here bear any responsibility for any analyses or conclusions.

References

- Backstrom, L., Sun, E., & Marlow, C. (2010). Find me if you can: Improving geographical prediction with social and spatial proximity. *Proceedings of the 19th International Conference on World Wide Web* (pp. 61–70). doi: 10.1145/1772690.1772698
- Cranshaw, J., Schwartz, R., Hong, J. I., & Sadeh, N. (2012). The livelihoods project: Utilizing social media to understand the dynamics of a city. *Proceedings of the Sixth International AAAI Conference on Weblogs and Social Media* (pp. 58–65).
- El Esawey, M., Lim, C., Sayed, T., & Mosa, A. I. (2013). Development of daily adjustment factors for bicycle traffic. *Journal of Transportation Engineering*, 139(8), 859–871.
- Federal Highway Administration (FHWA). (2013). *Traffic monitoring guide: Chapter 4 traffic monitoring for non-motorized traffic VI*. Washington DC: U.S. Department of Transportation, Federal Highway Administration.
- Griffin, G., Nordback, K., Gotschi, T., Stolz, E., & Kothuri S. (2014). Monitoring bicycle and pedestrian travel behavior: Current research and practice. *Transportation Research Circular E-C183*. Washington, DC: Transportation Research Board of the National Academies.
- Griffin, G. P., & Jiao, J. (2015). Where does bicycling for health happen? Analyzing volunteered geographic information through place and plexus. *Journal of Transport and Health*, 2(2), 238–247.
- Hankey, S., Lindsey, G., & Marshall, J. (2014). Day-of-year scaling factors and design considerations for non-motorized traffic monitoring programs. *Transportation Research Record*, 2468, 64–73.
- Jackson, K., Stolz, E., & Cunningham, C. (2015). Non-motorized site selection methods for continuous and short-duration volume counting. *Transportation Research Record*, 2527, 49–57.
- Jestico, B., Nelson, T., & Winters, M. (2016). Mapping ridership using crowd-sourced cycling data. *Journal of Transport Geography*, 52, 90–97.
- Keeler, B. L., Wood, S. A., Polasky, S., Kling, C., Filstrup, C. T., & Downing, J. A. (2015). Recreational demand for clean water: Evidence from geotagged photographs by visitors to lakes. *Frontiers in Ecology and the Environment*, 13(2), 76–81.
- Lenormand, M., Tugores, A., Colet, P., & Ramasco, J. J. (2014). Tweets on the road. *PLOS ONE*, 9(8), e105407.
- Lindsey, G., Petesch, M., & Hankey, S. (2015). The Minnesota bicycle and pedestrian counting initiative: Implementation study. MnDOT Report No. 2015-34. Minneapolis, MN: Minnesota Department of Transportation, Office of Transit, Bicycle/Pedestrian Section.
- Lindsey, G., Wilson, J., Rubchinskaya, E., Yang, J., & Han, Y. (2007). Estimating urban trail traffic: Methods for existing and proposed trails. *Landscape and Urban Planning*, 81(4), 299–315.
- Metropolitan Council. (2016). *Annual Use Estimate of the Metropolitan Regional Parks System for 2014*. St. Paul, MN: Metropolitan Council. Retrieved from <http://metro council.org/Parks/Publications-And-Resources/PARK-USE-REPORTS/2014-Annual-Use-Estimate-of-the-Regional-Parks-Sys.aspx>
- Miguéns, J., Baggio, R., & Costa, C. (2008). Social media and tourism destinations: TripAdvisor case study. *Advances in Tourism Research*, 26(28), 1–6.
- Minge, E., Falero, C., Lindsey, G., & Petesch, M. (2015). *Bicycle and pedestrian data collection manual (DRAFT)*. MnDOT Report No. MN/RC 2015-33. Minneapolis, MN: Minnesota Department of Transportation, Office of Transit, Bicycle/Pedestrian Section.
- Miranda-Moreno, L. F., Nosal, T., Schneider, R. J., and Proulx, F. (2013). Classification of bicycle patterns in five North American cities. *Transportation Research Record*, 2339, 68–79. doi: 10.3141/2339-08

- Nordback, K., Marshall, W., Janson, B., & Stolz, E. (2013). Estimating annual average daily bicyclists: Error and accuracy. *Transportation Research Record*, 2339, 90–97.
- Richards, D. R., & Friess, D. A. (2015). A rapid indicator of cultural ecosystem service usage at a fine spatial scale: Content analysis of social media photographs. *Ecological Indicators*, 53, 187–195.
- Sessions, C., Wood, S. A., Rabotyagov, S., & Fisher, D. (2016). Measuring recreational visitation at U.S. national parks with crowd-sourced photographs. *Journal of Environmental Management*, 183, 703–711.
- Sharp, R., Tallis, H. T., Ricketts, T., Guerry, A. D., Wood, S. A., Chaplin-Kramer, R., ... Bierbower, W. (2016). *InVEST +VERSION+ User's Guide*. The Natural Capital Project, Stanford University, University of Minnesota, The Nature Conservancy, and the World Wildlife Fund.
- Verma, R., Stock, D., & McCarthy, L. (2012). Customer preferences for online, social media, and mobile innovations in the hospitality industry. *Cornell Hospitality Quarterly*, 53(3), 183–186.
- Wang J., Hankey, S., Wu, X., & Lindsey, G. (2016). Monitoring and modeling urban trail traffic: Validation of direct demand models in Minneapolis, Minnesota, and Columbus, Ohio. *Transportation Research Record*, 2593, 47–59. doi: 10.3141/2593-06
- Wang, X., Lindsey, G., Hankey, S., & Hoff, K. (2014). Estimating mixed-mode urban trail traffic using negative binomial regression models. *Journal of Urban Planning and Development*, 140(1).
- Wood, S. A., Guerry, A. D., Silver, J. M., & Lacayo, M. (2013). Using social media to quantify nature-based tourism and recreation. *Scientific Reports*, 3, 2976. doi:10.1038/srep02976