EXPLORING THE USE OF EGOCENTRIC ONLINE SOCIAL NETWORK DATA TO CHARACTERIZE INDIVIDUAL AIR TRAVEL BEHAVIOR

Thomas A. Wall[1] (corresponding author)
Doctoral Candidate
School of Civil & Environmental Engineering
Georgia Institute of Technology
790 Atlantic Drive
Atlanta, GA 30332-0355
Email: twall@gatech.edu
Phone: 404-894-2236
Fax: 404-894-5418

Gregory S. Macfarlane
Doctoral Candidate
School of Civil & Environmental Engineering
Georgia Institute of Technology
790 Atlantic Drive
Atlanta, GA 30332-0355
Email: gregmacfarlane@gatech.edu
Phone: 404-894-2236
Fax: 404-894-5418

Kari Edison Watkins, Ph.D.
Assistant Professor
School of Civil & Environmental Engineering
Georgia Institute of Technology
790 Atlantic Drive
Atlanta, GA 30332-0355
Email: kari.watkins@ce.gatech.edu
Phone: 404-385-4213
Fax: 404-894-2278

Original Submission Date: July 30, 2013
Revision Submission Date: September 1, 2013

Word Count: 5793 (text) + 0 (0 figures) + 1500 (6 tables) + 500 (≤ 35 citations) = 7793

[1] ACRP Graduate Research Award recipient

ABSTRACT
The rapid growth of online social networking over the past decade has generated tremendous amounts of data about individuals and their social relationships. Recent research studies investigating social relationships and travel behavior have sought connections between individuals' social networks and social-related travel; however, to our knowledge none have pursued the use of online social networking data to do so. In this study, we explore the use of online social network data in characterizing individuals' air travel behavior. Data are collected using a web-based survey that gathers information about individuals' air travel history and online social network information, specifically participants' Facebook networks. The data are then analyzed to address a series of hypotheses about the association between online social network characteristics (specifically Facebook) and air travel behavior; in particular, travel distance, leisure-related travel, and trip generation. This study finds that there is a positive relationship between the size and distribution of individuals' Facebook social networks and their engagement in air travel, and also the odds that their air travel will be leisure-related, or include a leisure component.

INTRODUCTION

Over the past several decades, transportation researchers have begun to investigate what has been called "the link between social networks, locational choices, and travel." *(1)* Of particular interest has been the possible link between individuals' social networks and travel behavior. However, despite the growing interest among researchers in this area, it has been noted that relatively little research has investigated the link between social networks and travel *(2-4)*.

The rapid growth of online social networking sites (e.g. Facebook) over the past decade has generated tremendous amounts of centralized data about individuals and their social relationships. As current online networking and social media technologies continue to grow, and as new technologies emerge, the possibility for applications of social network data in travel and travel behavior research is increasing rapidly.

The central purpose of this study is to explore the use of online social network data in characterizing individuals' travel behavior given the size and distribution of their online social networks. However, despite this study's exploratory nature, some empirical analyses are undertaken to investigate these applications. Specifically, the correlations between online social network characteristics and travel distance, leisure-related travel, and trip generation are investigated using individuals' online social network data and air travel history. Data are collected using a multi-part, web-based survey that gathers information about individuals' air travel behavior and their egocentric online social networks – focusing specifically on individuals' Facebook networks.

This paper begins with a brief background discussion of social travel, the connection with leisure travel, social network and travel research, and finally a characterization of online social networking sites. This is followed by a discussion of the development of the web-based survey instrument, characterization of the data collected, and a discussion of the statistical models used to evaluate the data. The statistical evaluation address three hypotheses related to online social networks and air travel: that individuals travel further to cities where they have friends than to cities where they do not; that individuals are more likely to take personal or leisure trips to cities where they have friends than to cities where they do not; and that individuals with more distant friends take more trips than individuals with more proximate friends. The paper concludes with a discussion of the model results, some thoughts on the applicability of this study's findings to current practices, and some considerations for future research in this area.

THE SOCIAL DIMENSION AND LEISURE TRAVEL

Traditionally, the study of travel behavior has focused on trip-based approaches (e.g. the four-step travel demand model), and more recently on activity-based approaches *(5, 6)*. It has been noted with respect to this latter method that activities are in fact, frequently *joint* activities involving family members, household members, or social network connections *(5)*. It has been further suggested that the spatial aspects of one's social connections may induce, or in other ways determine (i.e. alter or constrain) travel behavior *(5, 7)*. Given these assertions, as well as the general increase in travel for social purposes *(8)*, there is a strong need to investigate this "social dimension" *(9)* of travel. Doing so will complement the currently individualistic, activity-based approach, and uncover ways to better account for the number of individuals participating in an activity and the relationships among them *(10)*. However, despite this need, surprisingly little research has focused on the connection between social relationships and travel *(1, 3, 4)*.

Although even mandatory and maintenance trips can have the social elements described above, there is clearly a strong connection between the social dimension of travel and leisure trips. Citing Larsen et al. *(11)*, Carrasco and Miller *(9)* suggest that social and leisure activities have become central in individuals' lives, and that social-related travel is an essential component in both of these activities. As disposable incomes have increased and the real price of travel has decreased in the Western world *(12)*, there is certainly greater opportunity for more social-related leisure travel. This is perhaps evident in the observation that leisure travel is "the fastest growing segment of travel in terms of the share of trips and the share of miles traveled," and survey evidence has shown that such travel is predominantly social travel *(3)*.

SOCIAL NETWORKS

Social relationships among individuals represent the "potential [for] activity and travel between them." *(10)* Accordingly, much of the research into the social dimension of travel has focused on the analysis of social relationships in social networks (this research is discussed in the next section). Social networks may be viewed as conceptualizations of the *social relations* between individual actors or between groups of individuals. The two primary "dimensions" of social relations are their *structure* and their *function (13)*. As Due et al. *(14)* elaborates, the structure refers to "the individuals with whom one has an interpersonal relationship and the linkages between these individuals." These linkages include formal relations (e.g. professional relations) and informal relations (e.g. "linkages between individuals with whom one has a close family relation and/or affection," such as family, friends, and close colleagues) *(14)*. The *function* of social relations then refers to the "interpersonal interactions within the structure of the social relations;" for example, "social support, relational strain, or social anchorage." *(14)*

The essential components of a social network structure are the actors and the relationships among those actors. Social network structure can be studied either as whole networks or as egocentric networks *(15)*. Whole networks consist of all actors and relationships within an entire, defined population *(15, 16)*. Depending on the scale of analysis, whole network analysis for travel studies may be impractical. Egocentric networks, conversely, focus on the "set of ties surrounding sampled individual units" *(16)* – this central person is called the "ego" and those individuals with whom the ego has a relationship are called "alters." *(4)* By considering multiple egos, egocentric network analysis becomes the study of a sample of individuals from a population *(16)*, which may be more feasible when studying travel *(15)*.

SOCIAL NETWORK TRAVEL RESEARCH AND DATA COLLECTION

Research into social networks and travel has, to date, primarily focused on regional and urban mobility, and particularly through the study of information and communication technology (ICT) networks [for example, *(4, 17)*; and the *Connected Lives* study *(18)* and related travel studies *(9, 10, 15, 19, 20)*]. This also makes sense with respect to leisure activities, as it has been a noted that ICT (e.g. telephone, email, SMS) may "enable," or "facilitate" leisure activities *(21)*.

Data collection for social travel studies, however, can be a difficult and involved process. One challenge is that it is difficult for individuals to recall their own social network connections *(19)*. Thus, mechanisms such as name generators (specific questions designed to elicit some portion of the individuals' connections, particularly stronger ties, and to determine tie strength), are used to prompt individuals' responses *(19)* in mail-in questionnaires [for example, van den Berg et al. *(4)*] and in-person interviews. This latter method, however, can be time-intensive. For example, the *Connected Lives* study *(18)* used name generators in a series of face-to-face interviews to construct egocentric social networks for 87 individuals; these interviews took on average 2-1/2 hours each to complete *(19)*.

This study suggests that the use of online social networking sites may provide a more efficient means of collecting social network data for social travel research applications. Online social networking sites centralize and maintain extensive information about online egocentric social networks in a common structure, which is therefore not subject to a respondent's ability to recall social relationships. Some aspects of online social networks are discussed in the next section.

ONLINE SOCIAL NETWORKING

Membership in online social networking sites (SNS) has seen tremendous growth since their advent nearly a decade ago. At the end of 2012, 67% of adult internet users were found to use some form of SNS *(22)* [an increase from 61% in 2011 and 29% in 2008 *(23)*]. SNS usage now encompasses over 50% of all adults *(23)*, but varies significantly across demographics, perhaps most notably across age groups. Use of SNS among internet users in 2012 was found to be 81% among teens *(24)* [compared to 73% in 2009, and 55% in 2006 *(25)*]; 83% among young adults aged 18-29 years; 77% among users aged 30-49 years, and 52% among users aged 50-64 years *(22)*. Among the 65 and older population of internet users [over half of whom are now online *(26)*], SNS usage has also seen a strong increase from 13% in April 2009 *(26)* to 32% at the end of 2012 *(22)*.

With a global membership of 1.11 billion monthly active members *(27)*, Facebook is the most commonly used online social networking site *(25)*. Sixty-seven percent of all adult internet users are Facebook members *(22)*. Facebook user profiles typically include general demographic, educational, and geographic information. As users connect with other individuals, or "friends," this information is then shared across that relationship to create an online model of relationships and connections. The majority of adult SNS users in the United States express that the primary motivation for using such sites is to maintain current relationships (i.e. friends and family members), or to reconnect former relationships (i.e. old friends) *(28)*. Thus, it may be reasonable to suspect that an individual's online social network structure represents, to some extent, an online model of that individual's real-life social network structure.

From a practical perspective, Facebook also allows for programmatic interaction with its site via the Facebook application programming interface (API). The Facebook API is commonly used to develop games or other so-called "apps" that integrate or interact with Facebook, but is also a powerful way of enabling users to grant third parties access to information associated with their accounts for research, promotional, or other purposes. The Facebook API is used extensively in this study to access participants' online social networking information via a web-based survey.

FACEBOOK NETWORK AND AIR TRAVEL STUDY

To explore the use of online social networking data in travel behavior research, this study conducts an empirical analysis of individuals' recent air travel history and their Facebook networks to address three specific hypotheses:

1. Individuals tend to travel further to reach destinations where they have friends than to reach destinations where they do not have friends;
2. Individuals are more likely to take personal (i.e. non-business) trips, or to at least introduce a personal or leisure component to trips that are otherwise business-focused, to destinations where they have friends than to destinations where they do not; and,
3. Individuals with more distant friends travel more, in general, than individuals with more proximate friends.

To investigate these hypotheses, data were collected from a predominantly student population using a web-based survey. The data were then analyzed using a series of statistical models specific to each hypothesis. This section discusses the development of the web-based survey instrument, characterizes the data collected, and presents the statistical models and their results.

Web-Based Survey Instrument Design

Data were collected using a web-based survey instrument developed specifically for this study, which collects data in two ways. First, a two-part user-response section asks participants to enter various demographic and travel information. Next, an automated portion asks participants to log in to their Facebook accounts and then collects information from the user's profile and friends list via the Facebook application programming interface (API). Survey participants are informed of the data types collected at the beginning of the survey and consent is required before they are allowed to proceed.

The data collected during the user-response section consists of two types: general socio-demographic data, and a travel diary. The socio-demographic data items are shown in Table 1. All data fields are required for participants to proceed to the next survey step. Survey participants are then asked about air travel that they have completed in the 12 months prior to taking the survey. The survey suggests that participants consult any airline frequent flier accounts they maintain in a separate internet browser tab to aid in recalling trip details. The airline trip data items collected are also shown Table 1. Note that air travel in this study is constrained to origins and destinations in the United States and Canada. Also note the final two data items, "Other Major City Visited..." These fields are only made available to participants when the "Trip Type" is indicated to be a "Multi-Destination Trip." The fields "Origin Airport Code" and "Destination Airport Code" are searchable fields that allow participants to partially enter city names or airport names, and then choose from a self-populating list; associated state and country fields are then auto-filled.

Upon completing the user-response portion of the survey, participants then interact with the automated data-collection portion. This portion executes a web-based application via the Facebook API that collects anonymous information from the survey participant's Facebook profile and friends list. Participants are first informed of the data items that are collected via the Facebook API, as well as the restrictions of the application (these are discussed later). Participants then log in to their Facebook

TABLE 1 Socio-Demographic and Air Travel Survey Data Items and Value Ranges

Socio-Demographic Data Item	Possible Value Ranges
User ID	Unique numerical identifier
Age	18 to 100 years
Gender	Male / Female
Race	Asian
	African American
	Caucasian
	Hispanic
	Other
US Citizenship	Yes / No
Occupation	Unemployed
	Student
	Employed, Part-Time
	Employed, Full-Time
	Retired
Education (Highest Completed)	None
	High School
	Some College
	Bachelors Degree
	Masters Degree
	Ph.D. Degree
Income	$0 to $250,000 (in $10,000 increments); and >$250k
Air Travel Data Item	**Possible Values/Range**
Trip ID	Unique numerical identifier
User ID	Unique numerical identifier
Departure Date	Month/Day/Year
Trip Length	0 to 14 nights
Trip Type	Round-Trip
	One-Way Trip
	Multi-Destination Trip
Trip Purpose	Business
	Personal
	Both
	Other
Origin Airport Code	Airport Code
Origin State	State/Province
Origin Country	US, Canada
Destination Airport Code	Airport Code
Destination State	State/Province
Destination Country	US, Canada
Other Major City Visited 1	Airport Code
Other Major City Visited 2	Airport Code

accounts to complete the survey. This login procedure is handled by Facebook, not by the web-based survey site: participants select a button on the survey screen that opens a "pop-up" window, which automatically redirects to the Facebook login interface. Participants then enter their login information directly into the Facebook login site and, upon successfully logging in, the "pop-up" window closes, enabling participants to re-access the survey website. Participants are then given the choice to complete the survey (allowing collection of their Facebook profile and friend data), or to quit the survey and not have any Facebook data collected. If participants choose to complete the survey, their data are collected to a secure server at Georgia Tech, and they are then automatically logged out of Facebook. If participants choose to quit, they are automatically logged out of Facebook and no Facebook data are collected. For security purposes, if a participant closes their internet browser, or browser tab, after having successfully logged in to Facebook (thus terminating the survey), they are automatically logged out of Facebook. The data collected by the automated portion of the survey are shown in Table 2.

TABLE 2 Facebook User Profile and Friend List Data Items and Value Ranges

User Data Item	**Possible Value/Range**
Birth date	(Month/Day/Year), (Month/Day), or NULL
Survey Completion Timestamp	Time, Month, Day, Year
Completion Indicator	Binary indicator
Friend List Size (Total)	Number
Friend List Size (Collected)	Number
Friend Data Item	**Possible Value/Range**
Friend ID	Unique numerical identifier of friend (i.e. network alter)
User ID	Unique numerical identifier of user (i.e. network ego)
Hometown Location	(City), (City, State), (City, State, Country), NULL
Current City Location	(City), (City, State), (City, State, Country), NULL
Birth Date	(Month/Day/Year), (Month/Day), or NULL

Data collected during the automated portion of the survey are subject to the account settings and privacy restrictions of the participant's Facebook account, and to those of their friends' Facebook accounts. For example, consider an individual who has entered her "hometown location" and "current city location" into her Facebook profile, but has set her privacy settings such that her "hometown location" is not accessible to some sub-group of her connections. If a survey participant is friends with this individual, but is within the sub-group of friends that does not have access to the friend's hometown location, then the automated survey application is only able to access the current city information. In other words, the automated survey application can only access data that are available to the current survey participant's Facebook account. Additionally, it should be noted that users frequently enter incomplete information into certain data fields in their Facebook profile; for example, birth dates frequently omit the birth year, or are not entered at all.

Survey Response and Descriptive Statistics

Survey participants were recruited from the faculty and student populations within the School of Civil & Environmental Engineering at Georgia Tech. Undergraduate and graduate students were recruited by visiting common classes during the Spring 2013 and Summer 2013 semesters to inform potential participants of the survey (a follow-up email was then sent with the web-based survey URL). Faculty and additional graduate students were recruited using departmental email lists. Given some overlap in the invitation groups, it is estimated that 300-400 individuals received invitations to participate.

In total, 89 individuals participated in the study. Thirty-six of these participants did not complete the final step of entering their Facebook login information and allowing the automated survey application to access their Facebook account. Two participants who did complete the final step (i.e. allowed access to their Facebook account) had individual privacy settings within their Facebook accounts that prohibited the automated survey application from fully accessing their profile and friends list. Therefore, complete data sets containing demographic, travel, and Facebook user profile and friends list data were collected for 51 participants (an approximate 15% response rate given the assumed invitation population). Table 3 shows descriptive statistics of the User, Friend, and Trip data collected.

TABLE 3 Descriptive Statistics of Survey Participant Data

User Data		**Trip Data**	
Respondents (Total)	**89**	**Total Trips (All Types)**	**197**
Respondents with Facebook (All)	53	Round Trips	159
Respondents with Facebook (OK Privacy)	51	Multi-Destination Trips	9
Oldest Respondent (years)	61	One-Way Trips	29
Youngest Respondent (years)	19	**Average Trips Per Respondent**	**3.86**
Average Age (years)	**25.6**	**Trip Purpose Types**	
Median Age (years)	**24**	Business	68
		Personal	113
Friend Data		Both	14
Total Collected	**22,985**	Other	2
Hometown Provided	18,341	**Unique Airports Visited**	**50**
Current City Provided	21,178	US Airports	47
With at Least One U.S. Location	19,066	Canadian Airports	3
Number of Unique U.S. Locations	**3,347**	**Individual Destinations**	**200**

Geographic Data Processing

The data collected from the web-based survey were post-processed using ArcGIS 10.1® to enable further analyses of the data. Due to the Atlanta-based survey participant population, most trips were found to either originate or end in Atlanta, GA. Thus, the trip data were processed into destination data (e.g. individual multi-destination trips between Atlanta and other locations were broken into individual destinations). Destination airports were geo-located in ArcGIS to associate them with nearby metropolitan statistical areas (MSAs) by analyzing a 50 mile radius around the airport, and associating those MSAs that were found to overlap (partially or fully) with the radius area. The MSA 2010 census populations were used to define the potential travelshed associated with each airport. In many cases, airports were associated with more than one MSA, and vice versa (particularly in the Northeastern and Mid-Atlantic regions), reflecting overlapping travelsheds and the availability of multiple airport options to various areas.

Friend data were processed by first geo-locating hometown and current city locations as point locations. These point locations were then associated with nearby airports by analyzing a 50 mile radius around each destination airport to capture those locations that fell within the radius area. In many cases, current city or hometown locations were associated with multiple airports.

Statistical Modeling Methods

A series of statistical models were used to evaluate the individual hypotheses discussed earlier. The individual statistical modeling methods are discussed below in relation to the individual hypotheses being evaluated. The outcomes of these statistical models are discussed in the following sub-section.

Hypothesis One – Travel Distance and Social Networks

The first hypothesis is that people tend to travel further to reach destinations where they have friends than to reach destinations where they do not have friends. This hypothesis is tested with a linear regression model,

$$\boldsymbol{y} = X\boldsymbol{\beta} + \boldsymbol{\epsilon}$$

where $\boldsymbol{y}$ is a vector of length n containing the natural logarithm of the distance between Atlanta Hartsfield-Jackson International Airport and the destination airport. The attribute matrix X is an $n \times k$ matrix containing the $x_1 \dots x_k$ attributes of the trip, the trip-maker, or the trip-maker's social network (as described in Table 1 and 2). Our variable of interest is the number of friends the trip-maker has living within 50 miles of the destination airport; if the β coefficient associated with this variable is significantly positive, we can infer a positive correlation between the number of friends and the distance traveled. The stochastic error term $\boldsymbol{\epsilon}$ is assumed to be distributed independently and identically with a normal distribution. The elements $\beta_1 \dots \beta_k$ in the parameter vector $\boldsymbol{\beta}$ are the least-squares estimates for the marginal effect of X on $\boldsymbol{y}$; the semi-log specification means that the estimates represent the approximate percent change in $\boldsymbol{y}$**.**

Hypothesis Two – Leisure-Related Travel and Social Networks

The second hypothesis is that trips to destinations where the trip-maker has friends are more likely to be for personal reasons, or are more likely to include a leisure component, than to be entirely business-focused. This hypothesis is addressed with a binary logit model,

$$P(y_i = 1|X) = \frac{\exp(\boldsymbol{x}_i\boldsymbol{\beta})}{1 - \exp(\boldsymbol{x}_i\boldsymbol{\beta})}$$

where the probability of trip i including a non-business element is again a function of the trip and trip-maker attributes $x_{i1} \dots x_{ik}$ and the estimated parameters $\boldsymbol{\beta}$. The dependent variable is binary, with a value of one if the trip includes a personal component and zero if the trip is strictly for business purposes. The variable of primary interest is the number of friends the traveler has in the destination city; a positive coefficient indicates that people are more likely to take personal or leisure trips to cities where they have friends. In the case of the logit model the estimated parameters $\boldsymbol{\beta}$ represent the logarithm of the change in odds, rather than the direct marginal effect as above.

Hypothesis Three – Social Network Distribution and Increased Travel

The third hypothesis is that individuals with distant friends may actually travel more, all else constant, than those with proximate friends; in other words people whose social networks are away from Atlanta make more trips than people whose networks are local. There are many reasons to expect this to be so. Ostensibly, people with dispersed social networks have potentially more reasons to travel, or at least fewer reasons to remain in Atlanta on, for example, holiday weekends. The network dispersion may also be seen as an instrument for other unobservable or endogenous variables: it may be that people with dispersed social networks have acquired them through broad experience and frequent travel. From an econometric perspective, this hypothesis is best addressed with a count model, where the discrete number of trips η taken by individual i is a function of the conditional mean of an assumed density function. Because individuals who make zero trips are excluded from the sample, a zero-truncated assumed density distribution must be used. The model is,

$$P(y_i = \eta_i) = \frac{f(\eta_i)}{1 - f(0)}$$

where the assumed distribution $f(\eta)$ may be either a Poisson or a negative binomial distribution. The two distributions have the same conditional expectation, $\mathrm{E}[y_i|\boldsymbol{x}_i] = \exp(\boldsymbol{x}_i\boldsymbol{\beta})$. In these models, a unit increase in x_k leads to an approximately β_k-percent increase in the expected number of trips. The models differ in the variance: whereas the Poisson distribution has a mean equal to its variance $[\mathrm{Var}(y_i|\boldsymbol{x}_i) = \exp(\boldsymbol{x}_i\boldsymbol{\beta})]$, the negative binomial has an overdispersed variance,

$\text{Var}(y_i|\boldsymbol{x}_i) = \exp(\boldsymbol{x}_i\boldsymbol{\beta}) + \alpha \exp(\boldsymbol{x}_i\boldsymbol{\beta})^2$. The estimable parameter α, if statistically greater than 0, indicates the necessity of the negative binomial distribution. The estimated model parameters $\boldsymbol{\beta}$ represent the marginal effect of $\boldsymbol{x}$ on the expected value of y_i.

All three of the models used in the analysis assume that the errors are distributed independently, which may not always be the case in this study's data: several respondents made multiple trips to the same destination. Future research should apply mixed or random effects models to remove such within-group correlation.

Results

Hypothesis One – Travel Distance and Social Networks

Results from the linear regression models are given in Table 4. The "Base" model establishes the relationship between the trip-maker attributes and the distance traveled. The R^2 fit statistic is not unreasonable for a disaggregate behavior model [for example, see *(29, 30)*] but few of the predictor variables are significant. Men and holders of a graduate degree take moderately longer trips than women or people with a bachelor's degree or less. The population is also significant, with people traveling shorter distances to reach larger cities. This is somewhat unintuitive, as economic theory suggests that larger cities attract trips from further away; a likely explanation is that most large American cities are on the eastern seaboard, and therefore closer to Atlanta.

TABLE 4 Linear Models (ln(Distance Traveled))

	Base Model		**Current Friends**		**Hometown Friends**	
	β	t-*stat*	β	t-*stat*	β	*t-stat*
Intercept	10.709	11.78**	11.021	10.63**	11.530	11.61**
Age	0.000	0.02	0.000	-0.05	0.001	0.12
Male	0.220	2.58*	0.210	2.43*	0.202	2.38*
Minority	-0.066	-0.79	-0.059	-0.69	-0.028	-0.33
Student	-0.096	-0.68	-0.110	-0.78	-0.135	-0.96
Graduate Degree	0.164	1.74	0.162	1.71	0.137	1.45
ln(Income)	0.032	0.56	0.024	0.40	0.011	0.19
ln(Population)	-0.280	-6.05**	-0.293	-5.81**	-0.317	-6.39**
ln(Friends-Current)			0.007	0.63		
ln(Friends-Hometown)					0.017	1.98*
Degrees of Freedom	190		189		189	
R^2 adj.	0.177		0.175		0.190	
ln(L)	-143.2		-143.0		-141.2	
p: Likelihood Ratio			0.52		0.04	

** $p < 0.01$, * $p < 0.05$

Two models then introduce the effect that social networks have on distance traveled. The model "Current Friends" shows that a percentage increase in the number of friends currently in a city has no effect on the distance traveled to get there, and that including this variable does not improve model fit in any meaningful way. The model "Hometown Friends" indicates that having friends who list the destination city as their hometown (we interpret "hometown" as a birthplace, or place of childhood attachment), on the other hand, does have a statistically significant effect on distance traveled. On average, a 1% increase in the number of friends whose hometown is the destination city is associated with a 1.7% extension in the distance traveled (the estimated coefficients are elasticities). What this model likely uncovers is the hometown of the traveler himself, as a substantial portion of his friends will likely have the same hometown; this may be more pronounced given the younger average age of the sample. On one hand this revelation seems trivial, but on the other hand, this may be a mechanism whereby airlines could identify the hometowns of customers under limited information scenarios.

Hypothesis Two – Leisure-Related Travel and Social Networks

Results of the binary logit models of trip purpose are given in Table 5, and show roughly similar results to the linear models discussed above. Based on the results of the "Base" model, trips to distant and large cities are more likely to involve a personal component. Minority respondents are also more likely to take trips with personal components. But whereas the impact of friends above was somewhat equivocal, in this case the presence of either current or hometown friends at the destination significantly and substantially increases the probability of a trip involving a personal component (current and hometown friends are highly collinear, and cannot appear in the model together). As an illustration, having 20 more friends in the destination city improves the odds that a trip will have a personal component from 1:2 (probability of 0.677) to 1:2.35 (probability of 0.702). It is also worth noting that both coefficient estimates – for current city and hometown friends – are safely within the other's 95% confidence interval, meaning that we cannot reject that the effect may be the same.

TABLE 5 Binary Logit Models (Personal Trip vs. Business Trip)

	Base Model		**Current Friends**		**Hometown Friends**	
	β	t-*stat*	β	t-*stat*	β	*t-stat*
Intercept	-12.511	-2.35*	-3.788	-0.65	-4.679	-0.80
ln(Miles to ATL)	0.995	2.96**	1.005	2.91**	0.873	2.52*
Age	-0.058	-1.22	-0.084	-1.68	-0.054	-1.09
Male	0.212	0.56	-0.062	-0.15	0.092	0.23
Minority	0.751	1.94	1.051	2.52*	1.137	2.69**
Student	0.660	1.07	0.256	0.39	0.419	0.65
Graduate Degree	-0.251	-0.57	-0.276	-0.60	-0.439	-0.95
ln(Income)	-0.215	-0.82	-0.494	-1.72	-0.381	-1.41
ln(Population)	0.596	2.60**	0.280	1.14	0.270	1.09
ln(Friends-Current)			0.186	3.58**		
ln(Friends-Hometown)					0.147	3.48**
$\rho_c^2 : \rho_0^2$	0.126 : 0.177		0.180 : 0.228		0.176 : 0.224	
ln(L)	-113.00		-105.94		-106.47	
p: Likelihood Ratio against Base Model			0.0002		0.0003	

$\ln(L)_0 = -137.24, \ln(L)_c = -129.21$; ** $p < 0.01$, * $p < 0.05$

Hypothesis Three – Social Network Distribution and Increased Travel

Results from the count models, which estimate the effect that friends have on the expected number of trips taken, are shown in Table 6. The significance of the dispersion parameter α in all three models indicates that the negative binomial model is most appropriate for these data. The "Base" model reveals again that most available predictor variables are not significant. Wealthier individuals take somewhat more trips, as do respondents that hold a graduate degree. The second model, "Friends," introduces the total number of friends, irrespective of location, into the specification. This model is somewhat more predictive as the $\ln(\mathcal{L})$ values indicate a statistically significant improvement in model fit. Furthermore, the new parameter is moderately significant, indicating that a 1% increase in the number of total friends increases the expected number of airline trips by approximately 0.463%. The final model, "Non-ATL Friends," considers whether those friends are in Atlanta or not. This variable is a much better predictor of trip-making behavior, being strongly significant and showing that a 1% increase in the number of friends away from Atlanta increases the expected number of airline trips by 0.552%.

TABLE 6 Count Models (Number of Trips)

	Base		**Friends**		**Non-ATL Friends**	
	β	*t-stat*	β	*t-stat*	β	*t-stat*
Intercept	-2.615	-1.33	-5.363	-2.12	-5.409	-2.30
α	1.023	2.00*	1.164	2.29*	1.241	2.39*
Age	0.019	0.54	0.046	1.29	0.046	1.37
Male	-0.192	-0.67	-0.087	-0.31	-0.052	-0.19
Minority	0.049	0.17	0.149	0.53	0.122	0.44
Student	0.347	0.69	0.114	0.23	-0.008	-0.02
Graduate Degree	0.682	2.09*	0.561	1.78*	0.545	1.73*
ln(Income)	0.301	1.79*	0.229	1.39	0.192	1.16
ln(Total Friends)			0.463	1.78*		
ln(Friends away from ATL)					0.552	2.12**
ρ^2	0.223		0.235		0.241	
$\ln(\mathcal{L})$	-97.67		-96.21		-95.51	
p: Likelihood Ratio			0.088		0.038	

$\ln(\mathcal{L})_c = -136.29$; ** $p < 0.01$, * $p < 0.05$

DISCUSSION AND CONCLUSIONS

This study explores the use of online social network data to characterize individuals' air travel behavior given the size and distribution of their online social networks. Part of the motivation for this research has been the growing interest in the general importance of the social dimension as a relevant aspect of travel behavior, but also specifically the connection between socially-influenced travel and the rapidly growing leisure travel market. To investigate the usability of online social network data in travel behavior research, an empirical study was conducted that combines individuals' air travel diary and online social networking data (from the Facebook SNS) to address three hypotheses. These hypotheses are: (1) people tend to travel further to reach destinations where they have friends than destinations where they do not; (2) people are more likely to take personal trips, or to introduce a personal component to trips that are otherwise business-focused, to destinations where they have friends than to destinations where they do not; and (3) people who have more distant friends travel more than people who have more proximate friends.

The statistical analyses performed on the travel and online social network data collected exhibit several positive results that suggest online social network data can be used effectively to characterize aspects of individuals' air travel behavior. First, results from the linear models used to address the first hypothesis indicate that there is a significant positive relationship between the distance an individual is willing to travel to a destination and the number of connections who identify that destination as their hometown. As mentioned, it is very possible that this is actually an indication of the traveler's own hometown, as one might suspect that an individual would have a larger number of social network connections in their hometown. In that respect, this finding may be useful as a proxy for an individual's hometown when such information is not readily available.

Results from the binary logit models used to address the second hypothesis indicate a positive and significant relationship between the number of connections that a traveler has associated with a particular destination and the odds that a trip to that destination will either be for personal or leisure purposes, or have some personal or leisure component. The count models used to address the final hypothesis indicate that individuals with greater numbers of total online social network connections may engage in greater air travel. However, a more significant result is the indication that individuals with greater numbers of connections away from their current location may be expected to engage in greater air travel.

The empirical findings of this study have potential applications to airline marketing strategies, route planning, and to airport planning and policy. First, the findings that link an individuals' total number of Facebook friends, their number of distant (versus proximate) Facebook friends, and their expected number of air trips, can help to better characterize an individual's general propensity to engage in air travel. This information could be useful to airlines for travel marketing purposes, but also for route planning purposes. It could also be useful to airports in better characterizing the potential travelers who live within an airport's service area and their propensity for travel.

The connection between an individual's online social network and the odds that a trip will be personal or leisure-related (or will contain some component thereof) is particularly interesting. In this study, personal, touristic, and leisure travel are grouped under the common label "leisure" (or more generally, "non-business related" travel). While it may be easier to identify touristic travel destinations, identifying potential destinations or corridors for personal or non-touristic leisure travel (e.g. visiting friends and relations, or so-called "VFR" travel) can be difficult. The use of data from Facebook or other online social networking sites may help airlines in identifying potential leisure travel routes, and also help airports in identifying leisure travelers from among their travelshed. Additionally, such data could also be helpful in identifying *potential* leisure travelers among current business travelers. For example, it could be used to assist airlines in setting fares that encourage business travelers to extend certain itineraries as a means to enable or induce leisure activities.

This study constitutes an early effort to investigate the use of online social network data in travel behavior research. As such, several known limitations to this research should be addressed in future efforts. First, the small sample size of survey respondents and use of a largely Atlanta-based, student population limits the significance of the findings and the range of hypotheses that can be evaluated. Students are generally less able to afford air travel than higher-income individuals. Additionally, the Atlanta-centric nature of the study population only enables an evaluation of trips taken from Atlanta to other destinations, rather than a more thorough destination choice approach. Given the greater usage of Facebook (and other online SNS) among the 18-29 year demographic as compared to older demographics *(22)*, the individuals in this study's population have, on average, more Facebook friends than the typical user [the average friend list size in this study was found to be 617 friends, as compared to the 2011 global average of 190 friends *(31)*]. This has implications for the applicability of this study's finding in characterizing air travel behavior for older age groups. However, the efficacy of using online social networking data for travel behavior research may improve over time as it has been suggested that the current Millennial generation "will retain their willingness to share personal information online [in social networking sites] even as they get older." *(32)*

Next, the relationship between an individual's actual social network and their online social network should be assessed further. One assumption in this study has been that the structure of an individual's online social network provides a reasonable representation of his or her actual social network structure. While future research should seek to validate this, an equally important aspect is to characterize the *function* of relationships within an individuals' online social network. Tie-strength, for example, may be one aspect of this. Many of the social network and social travel studies discussed earlier incorporate mechanisms to assess tie-strength among social relations (e.g. name generators, name interpreters) [see, Carrasco et al. *(19)*]. However, social relationship tie-strength is difficult to determine from Facebook data exclusively. One implication of this is that online social networking data, such as Facebook data, may be a valuable complement to conventionally collected social network data. Nonetheless, future research should investigate novel approaches to assess tie-strength using the data that are accessible via the Facebook API.

One final general limitation is data availability. It is unclear how willing various social networking sites are to make data available for research purposes. Indeed, the survey-based nature of this research project was designed to address this challenge. However, if future research is to develop statistically significant results that are more broadly meaningful to the air travel industry, larger data samples (likely beyond the practical bounds of a survey-based study) will be necessary. For this reason, future research may benefit from partnerships among industry stake-holders (e.g. airlines, online travel agencies, airports) in pursuing data samples directly from social network sites.

Although this study has focused exclusively on the Facebook SNS, research opportunities exist among the myriad other social media and social networking sites [for a broad discussion, see *(33)*]. For example, the business networking site LinkedIn may provide insight into business travel behavior; the location-based social networking site Foursquare may be useful in studying urban or regional travel behavior; and other mobile social media services with location-based components (e.g. Twitter, Instagram) may contain data relevant to travel behavior. Given the fast pace of social media and mobile technologies, it will be necessary for travel researchers to keep abreast of emerging technologies over time. Nonetheless, the findings of this study suggest a rich future for the use of online social networking data in travel behavior research.

ACKNOWLEGEMENTS
The authors wish to thank the Airport Cooperative Research Program (ACRP) of the National Academies for sponsoring this project, and the National Center for Transportation System Productivity and Management (NCTSPM) at the Georgia Institute of Technology for its additional support. The authors wish to thank Neeraj Joshi and Lakshmi Neeraja Vinjam for their technical contributions, as well as Dr. Patricia Mokhtarian, and several additional reviewers, for providing helpful and constructive comments.

Thomas A. Wall received the ACRP Graduate Research Award in sponsorship of his development and implementation of this research project; Gregory S. Macfarlane provided invaluable statistical modeling expertise; Dr. Kari Edison Watkins provided expert guidance in overseeing this research project.

REFERENCES
1. Axhausen, K. W. Social Networks, Mobility Biographies, and Travel: Survey Challenges. *Environment and Planning B: Planning and Design*, Vol. 35, 2008, pp. 981-996.
2. Molin, E., T. Arentze and H. Timmermans. Social Activities and Travel Demand: Model-Based Analysis of Social-Network Data. *87th Annual Meeting of the Transportation Research Board*, 2008,
3. Axhausen, K. W. Social Networks and Travel: Some Hypotheses. *Arbeitsbericht Verkehrsund Raumplanung, Institut für Verkehrsplanung und Transportsysteme, ETH Zürich, Zürich*, Vol. 197, 2003, pp.
4. van den Berg, P., T. A. Arentze and H. J. P. Timmermans. Size and Composition of Ego-Centered Social Networks Adn Their Effects on Geographic Distance and Contact Frequency. *Transportation Research Record*, Vol. 2135, 2009, pp. 1-9.
5. Sharmeen, F., T. Arentze and H. Timmermans. Modelling the Dynamics between Social Networks and Activity-Travel Behavior: Literature Review and Research Agenda. *12th World Congress on Transportation Research*, 2010,
6. Ronald, N. *Modelling the Effects of Social Networks on Activity and Travel Behaviour*. Doctoral Dissertation Eindhoven, The Netherlands, 2012.
7. Ohnmacht, T. Social-Activity Travel" Do the 'Strong-Tie Relationships' of a Person Exist in the Same Community? The Case of Switzerland. *Environment and Planning A*, Vol. 41, 2009, pp. 3003-3022.
8. Schlich, R., S. Schonfelder, S. Hanson and K. Axhausen. Structures of Leisure Travel: Temporal and Spatial Variability. *Transport Reviews*, Vol. 24, No. 2, 2004, pp. 219-237.
9. Carrasco, J.-A. and E. J. Miller. The Social Dimension in Action: A Multilevel, Personal Networks Model of Social Activity Frequency between Individuals. *Transportation Research Part A*, Vol. 43, 2009, pp. 90-104.
10. Carrasco, J.-A., B. Hogan, B. Wellman and E. J. Miller. Agency in Social Activity Interactions: The Role of Social Networks in Time and Space. *Tijdschrift voor Economische en Sociale Geografie (Journal of Economic and Social Geography)*, Vol. 99, No. 5, 2008, pp. 562-583.

11. Larsen, J., J. Urry and K. W. Axhausen *Mobilities Networks Geographies*. Ashgate, Hampshire, England, 2006.
12. Larsen, J., K. W. Axhausen and J. Urry. Geographies of Social Networks: Meetings, Travel and Communications. *Mobilities*, Vol. 1, No. 2, 2006, pp. 261-283.
13. O'Reilly, P. Methodological Issues in Social Support and Social Network Research. *Social Science and Medicine*, Vol. 26, 1988, pp. 863-873.
14. Due, P., B. Holstein, R. Lund, J. Modvig and K. Avlund. Social Relations: Network, Support and Relational Strain. *Social Science and Medicine*, Vol. 48, 1999, pp. 661-677.
15. Carrasco, J.-A. and E. J. Miller. Exploring the Propensity to Perform Social Activities: A Social Network Approach. *Transportation*, Vol. 33, 2006, pp. 463-480.
16. Marsden, P. Network Data and Measurement. *Annual Review of Sociology*, Vol. 16, 1990, pp. 435-463.
17. Senbil, M. and R. Kitamura. Simultaneous Relationships between Telecommunications and Activities. *10th International Conference on Travel Behaviour Research*, 2003,
18. Wellman, B., B. Hogan, K. Berg, J. Boase, J.-A. Carrasco, R. Cote, J. Kayahara, T. Kennedy and P. Tran. Connected Lives: The Project. In *Networked Neighbourhoods*, Springer, London, United Kingdom, 2006.
19. Carrasco, J.-A., B. Hogan, B. Wellman and E. J. Miller. Collecting Social Network Data to Study Social Activity-Travel Behavior: An Egocentric Approach. *Environment and Planning B: Planning and Design*, Vol. 35, 2008, pp. 961-980.
20. Carrasco, J.-A., E. J. Miller and B. Wellman. How Far and with Whom Do People Socialize: Empirical Evidence About Distance between Social Network Members. *Transportation Research Record*, Vol. 2076, 2008, pp. 114-122.
21. Mokhtarian, P. L., I. Salomon and S. L. Handy. The Impacts of Ict on Leisure Activities and Travel: A Conceptual Exploration. *Transportation*, Vol. 33, 2006, pp. 263-289.
22. Duggan, M. and J. Brenner. *The Demographics of Social Media Users - 2012*. Pew Internet and American Life Project. Washington, DC, 2013.
23. Madden, M. and K. Zickuhr. *65% of Online Adults Use Social Networking Sites*. Pew Internet and American Life Project. Washington, DC, 2011.
24. Madden, M., A. Lenhart, S. Cortesi, U. Gasser, M. Duggan, A. Smith and M. Beaton. *Teens, Social Media, and Privacy*. Pew Internet and American Life Project. Washington, DC, 2013.
25. Lenhart, A., K. Purcell, A. Smith and K. Zickuhr. *Social Media & Mobile Internet Use among Teens and Young Adults*. Pew Internet and American Life Project. Washington, DC, 2010.
26. Zickuhr, K. and M. Madden. *Older Adults and Internet Use*. Pew Internet and American Life Project. Washington, DC, 2012.
27. Facebook. Fact Sheet http://newsroom.fb.com/content/default.aspx?NewsAreaId=22 Accessed 4/24/2012.
28. Smith, A. *Why Americans Use Social Media*. Pew Internet and American Life Project. Washington, DC, 2011.
29. Bodea, T. D., L. A. Garrow, M. D. Meyer and C. L. Ross. Explaining Obesity with Urban Form: A Cautionary Tale. *Transportation*, Vol. 35, No. 2, 2008, pp. 179-199.
30. Adrian Saldarriaga-Isaza, C. and C. Vergara. Who Switches to Hybrids? A Study of a Fuel Conversion Program in Columbia. *Transportation Research Part A: Policy and Practice*, Vol. 43, No. 5, 2009, pp. 572-579.
31. Ugander, J., B. Karrer, L. Backstrom and C. Marlow. *The Anatomy of the Facebook Social Graph*. Ithica, NY, 2011.
32. Anderson, J. Q. and L. Rainie. *Millennials Will Make Online Sharing in Networks a Lifelong Habit*. Pew Internet and American Life Project. Washington, DC, 2010.
33. Bregman, S. and K. E. Watkins (Eds.) *Best Practices for Transportation Agency Use of Social Media*. CRC Press, New York, 2013.