**DEVELOPMENT OF A SOCIAL WEIGHTS MATRIX TO**

**CONSIDER FRIENDSHIP INFLUENCES ON AIR TRAVEL**

**Bingling Zhang**

Transportation Engineering Analyst

Kimley-Horn and Associates, Inc.

[office address]

Email: bing.zhang@kimley-horn.com

**Gregory S. Macfarlane, Ph.D.**

Engineer, Systems Analysis Group

Parsons Brinckerhoff

3340 Peachtree Road NE #2400

Atlanta, GA 30326

Email: macfarlaneg@pbworld.com

**Thomas A. Wall, Ph.D.**

Postdoctoral Researcher

School of Civil & Environmental Engineering

Georgia Institute of Technology

790 Atlantic Drive

Atlanta, GA 30332-0355

Email: twall@gatech.edu

**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

Original Submission Date: August 1st, 2014

**Word Count:** 5922(5922 text) + 500 (2 figures) + 500 (2 tables) + 1157 (41 citations) = 8079

**ABSTRACT**

Social networks can influence travel in two distinct ways: first, people often make trips to see people they know through work or life (exogenous effects); or they make trips that are similar to those taken by friends or family (endogenous effects). In spite of this, traditional models have not always made use of social network data, because information on social networks is both difficult to collect and challenging to implement in models. In this paper, we appropriate techniques developed for *spatial* data (specifically, a spatial Durbin count model) to construct a trip generation model for *socially* linked observations. We train this model on data collected through a web survey that interacts with the Facebook application programming interface (API) to capture respondents’ “friends” lists. We construct a set of “social weights” matrices that captures the social tie-strength between observations. A matrix that weights two observations by the number of their mutual friends provides the maximum model likelihood; this model further suggests that a non-social model may result in erroneous interpretations of the link between demographic characteristics and trip generation.

**INTRODUCTION**

Social networks are conceptualizations of an individual’s social connections, consisting of the people with whom an individual has relationships. An individual’s social network may include family, friends, peers, co-workers, and casual contacts *(1–3)*, each of whom may play different roles regarding activity participation *(4)*. Social networks have been studied in sociology for several decades and research suggests that an individual’s behavior is influenced by his friends *(4, 5)*.

Researchers are now considering how social networks influence individuals’ travel behavior. Such studies have focused primarily on the exogenous effects of social networks, suggesting that individuals tend to travel to see members of their social network *(6–13)*. However, there is limited research on endogenous effects, specifically the social interdependence of travel behavior (i.e., whether an individual has similar travel behavior as members of his or her social network). One reason for this may be that traditional social network data collection techniques are difficult and time-consuming *(14, 15)*. However, with an increasing number of people utilizing online social networking websites, a large dataset is amassing that highlights naturally occurring social networks *(16)*. Analysis of these datasets may be a more efficient means by which to gain insight into an individual’s social network as online data collection methods are less cumbersome than traditional methods.

To understand the social interdependence of travel behavior, it is necessary to quantify relationship intensity (i.e. tie-strength) in an individual’s social network *(17–22)*. Tie-strength can be determined by using frequency of contact by various modes, such as wall posts, messages, and common photos *(23–26)*. However, not all of this data is publically available.

The objective of this research is to develop and test a method for representing relationship tie-strength in an individual’s online social network based on limited data publically available from Facebook. This study uses existing econometric techniques to spatially analyze and consider the social interdependence of dependent variables. This paper begins with a review of the relevant literature regarding social influences on travel behavior, issues with traditional social network data collection methods, the use of online social networks data, and the tie-strength of relationships. The paper then discusses the econometric methodology selected, data collection and processing, and the development and testing of the candidate weight matrices which consider tie-strength.

**LITERATURE REVIEW**

**Social Influences on Travel Behavior**

Transportation is a derived demand wherein people make trips to perform activities. Research suggests that an individual’s travel behavior may be influenced by his social network *(6–8, 10, 13)*. One such study hypothesized that an individual’s social network is a key cause of travel behavior, finding that individuals who live closer together and more actively plan events are better able to maintain reciprocal relationships *(6)*. Additionally, individuals with larger or more fragmented social networks tend to be more active, as such networks require more maintenance *(8)*. Studies also indicate that individuals are willing to travel farther for and spend more time in social activities with members of their social network *(7, 10)*. A recent study by Wall et al (2014) suggests a positive relationship between an individual’s air travel and the size and geographic distribution of his online social network *(13)*.

**Traditional Social Network Data Collection**

There are two main approaches to analyzing the structure of social networks – egocentric network analysis and whole network analysis. The egocentric network approach focuses on a single individual (the ego) and his relationships with other people (alters). This approach is advantageous when the social network is large and has ambiguous boundaries. The whole network approach focuses on the relationships among all members in a network; this approach becomes difficult with large networks because the number of relationships increases rapidly *(1)*.

Name generators, sets of questions designed to solicit information from an individual regarding whom he considers as part of his social network, are frequently used by researchers to define social networks *(27–29)*. . However, problems with this data collection method can occur when respondents have difficulty in recalling information regarding their social networks *(15)*. Additional issues include when researchers attempt to infer whether the average behavior of a group influences the behavior of individuals; it is not possible to make inferences regarding exogenous effects, endogenous effects, or correlated effects (e.g. people with a common environment are likely to behave similarly) unless the researcher has previous knowledge relating to the composition of the reference group *(14)*. This type of data collection is also very time consuming, with the average interview taking over two hours *(30)*.

**Use of Online Social Networks Data**

Online social networking websites (e.g., Facebook and Twitter) provide an alternative method of social network data collection that may overcome some of the shortcomings of traditional methods. First, online social networks allow for naturally developing social networks with clear boundaries, thus overcoming respondent recollection issues *(14, 15)*. Second, as data can be retrieved from computer databases nearly instantly, data collection is less time consuming than traditional interview-based data collection methods *(14, 15)*.

Lewis (2008) discusses the five defining features of the Facebook dataset, which is used in this study. First, data from Facebook is naturally occurring when downloaded directly from Facebook.com, thus eliminating issues in a respondent’s ability to recall certain information *(14, 15)*. Second, Facebook.com contains complete network data of a sociocentric nature, as opposed to an egocentric nature, which reflects the interconnectedness of network members and member roles based on peers. Third, Facebook has the ability to provide longitudinal data, such as how relationships grow and change over time; however, such information may be unreliable as it requires active user input to terminated previously established relationships, which is rare. Fourth, Facebook can provide data on relationship strength. For example, people tagged together in photographs may have a stronger relationship than people who are only “friends”. Finally, Facebook data can provide some cultural insight as individuals can indicate tastes and values; however, this data is difficult to quantify and process as responses are generally open-ended. Additionally, privacy settings on Facebook often restrict the ability for researchers to retrieve certain data *(16)*.

**Relationship Tie-Strength**

Tie-strength is a measure of relationship intensity of in a social network and has been studied for both traditional and online social networks. Understanding and quantifying tie-strength has been researched in sociology and computer science, among other fields *(17–21, 23–26, 31)*. , and is typically assessed in traditional social network analysis during data collection by asking about communication frequency and the type of relationship.

Tie-strength research encompasses strong ties, intermediate ties, and weak ties, each of which has unique qualities *(18, 21, 22)*. Strong ties are associated with information propagation; however, a network with only strong ties may lead to fragmentation *(18, 21)*. Weak ties are more likely to link members of different social groups while strong ties are more likely to link people within social groups *(21)*. Research on intermediate ties is limited, but suggests that intermediate ties are important in inter-organizational knowledge transfer because they link key groups of organizations which are well connected with other organizations *(22)*.

Studies show that that the number of online social network relationships an individual can maintain is similar to the number of offline relationships maintained *(23)*. Additionally, Gilbert et al. (2009) compared tie-strength for individuals using offline and online methods and suggests that online social networks can be used to predict tie-strength *(24)*. . Thus, analyzing communication and interaction frequency in online social networks (e.g., mutual photos, wall posts, messages) may provide a reasonable assessment of tie-strength. . However, it should be noted that not all of this data is publically available from Facebook.

**METHODOLOGY**

**Econometric Methodology**

Econometric models have been used to show behavior similarity. For example, Bramoullé (2009) and Calvό-Armengol (2009) suggested that a student’s school performance and activity level is related to that of his friends *(32, 33)*. Spatial econometric models are traditionally used to show the spatial interdependence between variables. In transportation, for example, these models have been used to show the spatial effects of vehicle ownership *(34)*.

This study uses spatial econometrics, specifically the spatial Durbin model, to examine the social interdependence of air travel behavior. Spatial models are appropriate in this context as their weights matrices indicate the strength of the spatially related variables; in this study, the weights matrix represents individual social network tie-strength. The tie-strength weights matrices are row-standardized to create proportional relationships for each element, and are then tested in a spatial Durbin count model (social model) to determine the optimal weights matrix using maximum likelihood estimation. As a count model predicts the conditional mean based on a Poisson-distributed process, the social lag variables are considered by suggesting that . Previous literature on spatial count models include Bhat et al. (2013), which used a spatial Durbin count model to consider the relationship of several variables on the number of new business on a county level in Texas *(35)*. The social model is then compared to a non-social model to examine issues with socially dependent or correlated variable when using a non-social model be. A full treatment of spatial econometrics models is given by LeSage and Pace (2009).

**Data Collection**

. To examine the relationship between individuals’ social network and their air travel in interconnected social networks, researchers recruited participants from the Georgia Tech fraternity and sorority system in partnership with the Georgia Tech Dean of Students, Office of Greek Affairs. A web-based survey instrument hosted by the Georgia Institute of Technology’s School of Civil & Environmental Engineering was distributed via email to the primary points of contact for each fraternity and sorority on file at the Office of Greek Affairs. This primary sample population is approximately 3300 Georgia Tech undergraduate students, however fraternities and sororities were asked to also solicit responses from their alumni.. Fraternities and sororities were incentivized with a $500 contribution to a charity of their choice, made in their name, which was awarded to both the fraternity and the sorority with the highest number of survey respondents.

The online survey instrument was adapted from one developed for a previous survey *(13)*, and contains three components. For a detailed description of the variables collected, refer to Wall et al. (2014). The first component collected respondents’ air travel information from the past year – this includes a list of trips taken with origin-destination airports for round-trip, multi-destination, and one-way travel. The second component collected socio-demographic information, including age, gender, income, education level, citizenship, occupation, and race. The third component asked respondents to log in to their Facebook accounts, and the survey program automatically collected the hometown city and current city locations of the respondents, and also that of the respondent’s friends, via the Facebook application programming interface (API). Respondents were assigned a unique string identifier associated with their unique Facebook IDs using a one-way hash function (a common one-way encryption technique) to secure anonymity. This allowed researchers to identify any duplicate responses while maintaining individuals’ anonymity.

Data collection was limited by the individual account privacy settings of the respondent’s and respondent’s friends. For example, privacy settings prevented the collection of friendship, hometown location, and current city location data from certain individuals with more restrictive settings. Additionally, differing privacy settings between two friend connections may be reflected as one-way in the data even though one-way friend connections are not possible on Facebook.

**Data Preparation – Geospatial Processing**

Facebook hometown and current city locational data is not geocoded, and therefore requires some post-processing prior to any analyses. Facebook location data were post-processed using R statistical programming language *(36)* and ArcGIS 10 *(37)*.

The study analyzed respondents’ friends who do not consider Atlanta, Georgia as their hometown. Friends’ hometown cities were geocoded as point locations using ArcGIS 10 *(37)*. While most hometown cities were geocoded by matching the location names with a list of locations retrieved from a previous study *(13)*, researchers manually geocoded approximately 2% of cities that were not recorded previously. To associate cities as within (or outside of) the Atlanta metropolitan region, circular buffers were created (using Hartsfield-Jackson Atlanta International airport as the centroid) at 25 mile radius increments from 25 miles to 150 miles. R statistical programming language was then used to associate each of the 6 resulting hometown buffer regions with city locations, and thus determine the number of friends with hometown city locations within each buffer region. Only data from respondents that completed the entire survey (i.e. the air travel diary, socio-demographic section, and the Facebook data collection) were considered for the analysis.

**Candidate Weighting Schemes**

Four types of friend connections can be determined from the Facebook data collected in this survey: (1) direct friends only, (2) mutual friends only, (3) direct friends with mutual friends, and (4) no relationship. Direct friends only is defined as the two individuals being friends, as indicated by existing on each other’s friends list on Facebook, and having no other friends in common. For the purpose of this research, only survey respondents who are friends with other respondents and have no mutual friends in common are considered “direct friends only”. “Mutual friends only” is defined as two individuals that are not direct friends, but have friends in common. From the survey data, mutual friends are determined by a friend existing on the friends list for both survey respondents. For the purpose of this research, only first order mutual friends are considered. “Direct friends with mutual friends” is defined as two individuals who are direct friends and also have friends in common. “No relationship” is defined as two individuals that are neither direct friends nor have mutual friends.

Based upon these friendship characterizations, candidate weight matrices were developed and implemented in R statistical programming language to examine the tie-strength relationship between survey respondents. Based on the data collected and previous literature, four weighting schemes were considered: binary, number of mutual friends, percentage of mutual friends, and direct mutual relationships.

*Binary*

The binary weights matrix is adapted from basic spatial econometric weighting schemes *(38)*. The binary weights matrix only considers whether an individual is direct friends with another individual, and only considers the relationships between survey respondents. If two respondents are direct friends, their relationship is indicated by a value of one in the matrix; if the two respondents are not direct friends, their relationship is indicated by a value of zero in the matrix:

Ideally, the matrix should be symmetric prior to row standardization (i.e. if respondent A is friends with respondent B, then respondent B is also friends with respondent A). However, due to the possibility of asymmetric privacy settings on Facebook, the matrix was not symmetric. For example, if respondent A has more restrictive privacy settings than respondent B, then the survey instrument would be able to access data in respondent A’s Facebook account that indicates respondent A and respondent B being friends; conversely, when respondent B completes the survey, the survey instrument may not be able to access similar data indicating their friendship due to stricter privacy settings of respondent A. This privacy restriction affected approximately 10% of the relationships in the sample. However, symmetry was assumed for this matrix as Facebook does not distinguish relationship directionality; the binary matrix was manipulated to be symmetric. It is also notable that all survey respondents were friends with at least one other survey respondent.

*Number of Mutual Friends*

Existing literature discusses tie-strength as a function of mutual friendship *(24, 26)*. The number of mutual friends matrix considers the number of friends common to two respondents, regardless of whether or not the two respondents are direct friends. The number of mutual friends was determined by counting the number of unique friend Facebook IDs in common between two respondents’ friends lists.

Based on the dataset, this matrix was symmetric prior to row-standardization (i.e. respondent A’s mutual friends with respondent B are the same as respondent B’s mutual friends with respondent A). This supports that privacy settings may cause the asymmetry seen in the binary matrix, as symmetry in this matrix may be due to mutual friends’ privacy settings likely affect both respondents in the same manner. However, it should be noted that this may not be true for other datasets as individuals can classify friends into different groups on Facebook and apply different privacy settings for each group.

*Mutual Friend Percentage*

The mutual friend percentage matrix is adapted from a cohesiveness measure used in Adamic (2003) that considers a ratio of the connecting links divided by the total number of links possible *(39)*. The mutual friend percentage weights matrix considers only the number of friends the two respondents have in common. Mutual friend percentage is calculated as the total number of mutual friends divided by the sum of the total number of friends of both respondents.

This matrix is also symmetric prior to row-standardization for the same reasons as discussed above. However, as before, this symmetry may not always be observed in other datasets.

*Direct and Mutual Friend Percentage*

The direct and mutual friend percentage weights matrix is a combination of the binary matrix and the mutual friend percentage matrix, and considers whether or not the respondents are direct friends and also whether or not the respondents have mutual friends. This matrix adds a bonus to the mutual friends percentage matrix if the two respondents are themselves friends.

This matrix is not symmetric prior to row-standardization as the binary matrix is not symmetric (as discussed above). Similar to the binary matrix, symmetry was assumed and the matrix was manipulated to become symmetric.

**RESULTS**

**Survey Response**

A total of 92 individuals participated in the study; however, only 56 respondents completed all components of the survey. Descriptive statistics of the survey respondents are shown in TABLE 1.

Of the 56 survey respondents, 54 of the respondents were friends with at least one other respondent. FIGURE 1

TABLE Descriptive Statistics of Survey Responses

|  |  |
| --- | --- |
| **User Data** | |
| **Respondents (Total)** | **92** |
| Respondents with Facebook (All) | 56 |
| Respondents with Facebook (OK Privacy) | 56 |
| **Oldest Respondent (years)** | **55** |
| **Youngest Respondent (years)** | **18** |
| **Average Age (years)** | **20.6** |
| **Median Age (years)** | **20** |
|  |  |
| **Friend Data** | |
| **Total Collected** | **37,519** |
| Hometown Provided | 33,082 |
| Current City Provided | 35,499 |
| With at least One U.S. Location | 36,797 |
| **Number of Unique U.S. Locations** | **3264** |
| **Trip Data** | |
|
| **Total Trips (All Types)** | **97** |
| Round Trips | 76 |
| Multi-Destination Trips | 1 |
| One-Way Trips | 20 |
| **Average Trips per Respondent** | **1.73** |
| **Trip Purpose Types** |  |
| Business | 10 |
| Personal/Leisure | 76 |
| Both | 12 |
| **Unique Airports Visited** | **48** |
| U.S. Airports | 43 |
| Canadian Airports | 1 |
| International Airports | 4 |
| **Individual U.S. and Canadian Destinations** | **93** |

Of the 56 respondents, 44 had made air travel trips over the past year, with an average of 1.73 trips per respondent. Given the largely US-centric survey population, this study considers only travel to U.S. and Canadian destinations.

The survey collected friend connection information for 37,519 friends of the 56 survey respondents. The average number of friends per individual is 805 friends; the maximum and minimum number of friends per respondent was 1,995 friends and 209 friends, respectively. From the data collected, 36,797 friends indicated a U.S. location as either their hometown location or their current city location; of these locations, 3,264 were unique U.S. locations.

**Optimal Weights Matrix Selection**

Relevant variables are analyzed using a spatial Durbin count model. The dependent variable considered is the number of trips an individual takes within a one year period. Building upon a study by Wall et al. (2014), several demographic variables, as well as the number of friends an individual has who do not list cities proximate to Atlanta, Georgia as their hometown (determined using the six buffer regions previously described), were considered as independent variables. This choice was motivated by a study population that is largely undergraduate students and the assumption that the primary purpose of travel was to visit family and friends (i.e., business trips were excluded from the analysis). Additionally, Wall et al. (2014) found that there was a strong positive correlation between greater travel and more disperse (i.e. non-proximate) social networks *(13)*.

Each weight matrix was row-standardized and used to create lag variables (i.e. ) to be considered in the models. For each weights matrix, seven models were considered: a base model (including all dependent variables except number of friends not from Atlanta) and one model for each of number of friends not from Atlanta based on the six radii. The optimal matrix was determined by selecting that which produces the maximum log-likelihood.Model log-likelihoods are plotted in FIGURE 1. Each candidate weights matrix is indicated in the figure by a different colored line; the solid lines represent count models where the number of friends not from Atlanta is considered as absolute numbers, while the dotted lines represent count models where the number of friends not from Atlanta is considered as a percentage of the respondent’s total friends.

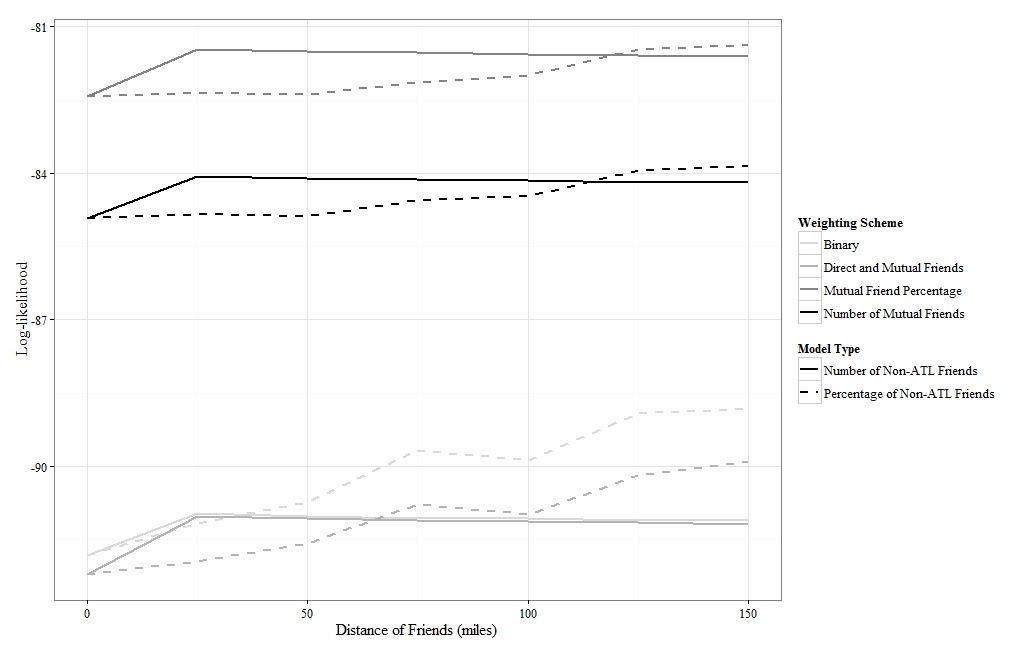


FIGURE Log-likelihoods for Candidate Weight Matrices Count Models

Optimal model likelihood occurs at a buffer distance of 25 miles using the mutual friend percentage weights matrix. Based on FIGURE 1, the mutual friend percentage weights matrix and the number of mutual friends weights matrix produce the highest log-likelihoods in the count models (both for absolute and percentage of non-Atlanta friends). This is interesting as it suggests that direct friendship is not as important as having mutual friends; both matrices which consider direct friendship have lower log-likelihoods. The binary matrix and the direct and mutual friends matrix have very similar log-likelihoods, which may be due to the binary component (indicating direct friendship) of the matrix influencing the model more than the mutual friend component.

FIGURE 1 also suggests that the six distance buffers used to consider the number of friends not from Atlanta have negligible effects on the log-likelihood of the model compared to the weight matrix selection. Using percentage of non-Atlanta friends as opposed to number of non-Atlanta friends produces higher log-likelihoods at larger buffer radii.

**Comparison of Social and Non-Social Model**

The optimal weights matrix used to characterize friendships is the mutual friend percentage matrix. The spatial Durbin count model (social model) results for this matrix are shown in TABLE 2 along with results of a model where the social influences are not considered (non-social model). It should be noted that the dataset is not a representative sample of the population; however, the impacts of considering social influences can still be seen in the results. Results show that, for this dataset, the non-social model and the social model indicate different variables as significant. While the non-social model indicates that an individual’s gender influences his trip count, the social model shows that it is actually the gender of his friends that has an influence on trip count. Similarly, the non-social model suggests that an individual’s income and non-Atlanta friends impact the number of air trips made annually. When social influences are considered, the model no longer shows those variables as being important. Instead, variables such as the individual’s race and his friends’ trip count and income show a strong influence on the number of air trips made by an individual.

TABLE Count Model Results

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **Non-Social Model** | | | **Social Model** | | |
|  | Estimates | Std. Error | | Estimates | Std. Error | |
| Intercept | 0.375 | 0.931 |  | 1.136 | 21.47 |  |
| Respondent's |  |  |  |  |  |  |
| Age | -0.040 | 0.028 |  | -0.155 | 0.033 | \*\*\* |
| Gender (M) | -0.658 | 0.255 | \*\* | -0.475 | 0.477 |  |
| Race (Minority) | -0.737 | 0.596 |  | -3.273 | 0.901 | \*\*\* |
| Log(Income) | 0.174 | 0.098 | . | 0.113 | 0.122 |  |
| Non-Atlanta Friends (25 Miles) | -6.40E-04 | 3.84E-04 | . | -0.001 | 4.43E-04 |  |
| Social Lags |  |  |  |  |  |  |
| Trip Count |  |  |  | -3.069 | 0.653 | \*\*\* |
| Age |  |  |  | 0.605 | 0.888 |  |
| Gender (M) |  |  |  | -6.678 | 0.397 | \*\*\* |
| Log(Income) |  |  |  | -0.166 | 1.644 | \*\*\* |
| Race (Minority) |  |  |  | -28.650 | 6.870 |  |

Significance Codes: ‘ . ’ 0.1 ‘ \* ’ 0.05 ‘ \*\* ’ 0.01 ‘ \*\*\* ’ 0.001

The research also explored the relationship between the frequency of an individual’s air trips compared to the frequency of his friends’ air trips. All four candidate weights matrices were considered in examining this relationship. The matrices which only consider mutual friendship (i.e. number of mutual friends, mutual friend percentage) and produce the highest log-likelihoods in the spatial Durbin count model show that an individual’s friends make slightly more air trips than the individual on average; matrices that consider direct friendship (i.e. binary, direct and mutual friends) indicate that that individual travels slightly more than his friends do.

FIGURE 2 shows the trip frequency relationships under the candidate weighting matrices as kernel density plots. The figure suggests that an individual’s friends make more trips than himself. This correlates to Feld’s friendship paradox, which states that there will always be someone more popular than the individual in his social network *(40)*. The study results suggest that in an individual’s social network, there will be someone who travels more.

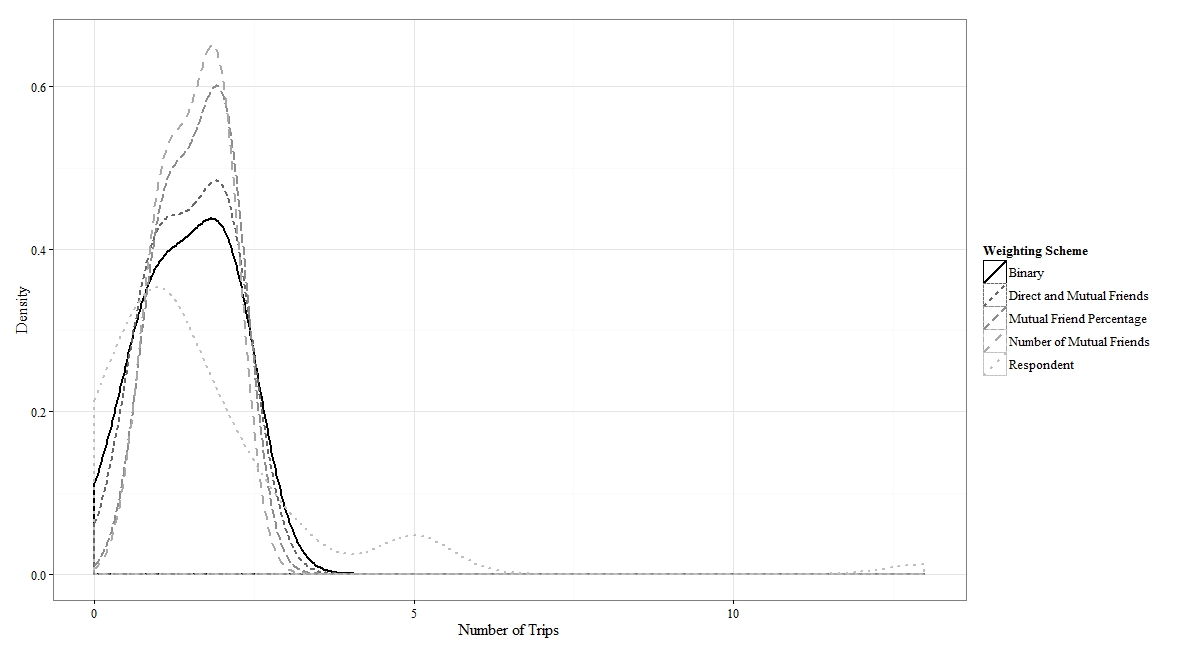


FIGURE Trip Frequencies of Friends by Weighting Scheme

**CONCLUSIONS**

This study explores the use of social tie-strength in characterizing the relationship between the air travel behavior of an individual and characteristics (including interdependence) of his online social network. Four different tie-strength weighting schemes were developed using data collected from an online survey that was distributed to the Greek system (fraternities and sororities) at the Georgia Institute of Technology. The survey instrument collected individuals’ demographic, air travel diary, and Facebook account friendship data. The candidate weight matrices were tested using a spatial Durbin count model, and relationships between individuals’ trip frequencies and social network characteristics were determined. Three major conclusions arise from this study.

First, out of the four candidate weighting schemes, the percentage of mutual friend weights matrix produces the highest log-likelihood in the spatial Durbin count model, based on the data collected. Additionally, the log-likelihoods suggest that the number of mutual friends that two individuals share is a more important indication of tie-strength than whether the two individuals are direct friends, as the two matrices that consider mutual friends exclusively produced higher log likelihoods than the matrices that also consider direct friendship. The results also show that if the optimal matrix is selected, the difference in the buffer radii distance used to consider the number of non-Atlanta friends is negligible in terms of model log-likelihood. This result suggests that based on the dataset, the percentage of mutual friends provides the best indication of the tie-strength between two individuals.

Second, the comparison of the non-social model with the social model indicates that if social influences are not taken into consideration, incorrect conclusions may be drawn. The non-social model suggests that an individual’s gender influences the number of air trips he takes annually, while the social model indicates that it is actually the gender of the individual’s friends that influences the trip count.

Third, the results suggest that an individual will have, on average, friends that make more trips than himself in his social network, which reflects the theory of Feld’s friendship paradox (there will always be someone more popular than the individual in his social network). This result suggests that an individual’s social network tends to include people who travel more than himself, indicating that people who travel more are likely to have a social network with other individuals who travel frequently.

**DATA LIMITATIONS**

The study is limited by the size of the dataset. Due to time constraints, the survey was released for two weeks at the end of the Spring 2014 semester to Georgia Institute of Technology students who were members of fraternities or sororities. The low response rate may be associated with the time frame in which responses were solicited; the end of the academic term is typically a busy period for students. The sample population is also not representative of the general public, as undergraduate students generally have lower incomes, which may lead to less air travel. These limitations may bias the dataset, as most respondents made fewer than five trips per year, and most of these trips were to the respondents’ hometowns.

Facebook policies also limited data collection. Information collected from Facebook accounts using the survey instrument was limited to the data that were accessible via the Facebook API, and thus also by individuals’ Facebook account privacy settings. These data only enabled the determination of one-degree removed friendships (i.e. friends of friends). Additionally, the survey did not collect other types of information (e.g. number of wall posts, mutually tagged pictures) that may be useful in determining tie-strength.

Furthermore, air travel history data collection is time-consuming and limited by respondents’ memory recollection. Although it was suggested to respondents that they reference frequent flyer account while taking the survey to aid in trip recollection, this portion of the survey could be automated, similar to the Facebook data collection portion, by retrieving data directly respondents’ frequent flyer accounts. This would eliminate memory recollection biases caused by the respondent and would be less time-consuming.

**FUTURE RESEARCH**

The contribution of this research is in determining methods of representing tie-strength using friendship data that is freely available from Facebook via the Facebook API. This research is a foundation for the consideration of social interconnectedness influences in travel behavior studies. Future research expanding on these contributions includes identifying different representations of tie-strength based on additional data and continued modeling of the social influence on travel behavior.

This study collected online social network data exclusively from Facebook.com to determine social tie-strength and examined only personal/leisure air travel to U.S. and Canadian destinations. Tie-strength determination can be expanded by collecting additional data from Facebook.com, such as number of wall posts or mutually tagged photos as discussed in Lewis et al. (2008). It could also expand to other online social networking sites (e.g. Twitter, Foursquare), or a combination of online social networks *(25)*. This information could be supplemented with more traditional in-person interviews in order to determine how accurately the online social network tie-strengths represent an individual’s offline social network tie-strengths. This research could also be expanded to include international air trips and business related air trips, as well as other transportation modes. Online social network websites could also be used in conjunction with traditional travel diaries to associate friendships with travel behavior.

Spatial models are currently used in transportation research primarily to represent how an individual’s behavior is similar to the behaviors of his neighbors, such as Paleti et al. (2012) which found that an individual’s vehicle type choice was influenced by his neighbors *(41)*. However, there has been little research as to whether this type of correlation is truly spatial or if the spatial correlation is actually a proxy for social correlation (i.e., where neighbors may also be social friends). Integrating online social network data into such spatial influence studies could help researchers distinguish between friends and neighbors, and thus social and spatial influences.

Lastly, this research could be further expanded to predict destination choice based on social networks. A dataset including social network data and air travel behavior, similar to the one used in this study, could be used in a discrete continuous model in order to predict the social influences on both the number of trips taken in a year and the distance traveled per trip.

Online social networks provide a wealth of data that can be used in conjunction with travel diaries to explore social influences on travel. As increasingly more people utilize online social networking websites, there is a growing dataset that is becoming available, which highlights naturally occurring social networks. Exploring these datasets can provide rich insight into an individual’s social network that is otherwise difficult and time-consuming to obtain through traditional data collection techniques. Utilizing these datasets to study how an individual’s air travel behavior is related to his online social network will help researchers better understand social activity-based travel and airline destination choice, which will help transportation professionals anticipate and address future transportation needs.

**REFERENCES**

1. Haythornthwaite, C. Social Network Analysis : An Approach and Technique for the Study of Information Exchange. No. 342, 1996, pp. 323–342.

2. Berg, P., T. Arentze, and H. Timmermans. A multilevel path analysis of contact frequency between social network members. *Journal of Geographical Systems*, Vol. 14, No. 2, Oct. 2010, pp. 125–141.

3. Marsden, P. V. NETWORK DATA AND MEASUREMENT. *Annual Review of Sociology*, Vol. 16, 1990, pp. 435–463.

4. Deutsch, K., and K. G. Goulias. Decision makers and socializers, social networks and the role of individuals as participants. *Transportation*, Vol. 40, No. 4, May 2013, pp. 755–771.

5. Viry, G. Residential mobility and the spatial dispersion of personal networks: Effects on social support. *Social Networks*, Vol. 34, No. 1, Jan. 2012, pp. 59–72.

6. 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. *Journal of Economic and Social Geography*, Vol. 99, No. 5, 2008, pp. 562–583.

7. 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: Policy and Practice*, Vol. 43, No. 1, Jan. 2009, pp. 90–104.

8. 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, No. 6, 2008, pp. 961–980.

9. Wellman, B., B. Hogan, K. Berg, J. Boase, J. Carrasco, R. Côté, J. Kayahara, T. L. M. Kennedy, P. Tran, N. Neighbourhoods, P. Purcell, and B. Springer. *Connected Lives: The Project*. 2005.

10. Moore, J., J.-A. Carrasco, and A. Tudela. Exploring the links between personal networks, time use, and the spatial distribution of social contacts. *Transportation*, Vol. 40, No. 4, Apr. 2013, pp. 773–788.

11. Sharmeen, F., T. Arentze, and H. Timmermans. An analysis of the dynamics of activity and travel needs in response to social network evolution and life-cycle events: A structural equation model. *Transportation Research Part A: Policy and Practice*, Vol. 59, Jan. 2014, pp. 159–171.

12. 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, No. 12, 2009, pp. 3003–3022.

13. Wall, T. A., G. S. Macfarlane, and K. E. Watkins. Exploring the Use of Egocentric Online Social Network Data to Characterize Individual Air Travel Behavior. *Transportation Research Record*, Vol. 2400, 2014, pp. 78–86.

14. Manski, C. Identification of Social Endogenous Effects : The Reflection Problem. *The Review of Economic Studies*, Vol. 60, No. 3, 1993, pp. 531–542.

15. Axhausen, K. W. Social networks, mobility biographies, and travel: survey challenges. *Environment and Planning B: Planning and Design*, Vol. 35, No. 6, 2008, pp. 981–996.

16. Lewis, K., J. Kaufman, M. Gonzalez, A. Wimmer, and N. Christakis. Tastes, ties, and time: A new social network dataset using Facebook.com. *Social Networks*, Vol. 30, No. 4, Oct. 2008, pp. 330–342.

17. Marsden, P. V., and K. E. Campbell. Measuring Tie Strength. *Social Forces*, Vol. 63, No. 2, 1984, pp. 482–501.

18. Shi, X., L. A. Adamic, and M. J. Strauss. Networks of strong ties. *Physica A: Statistical Mechanics and its Applications*, Vol. 378, No. 1, May 2007, pp. 33–47.

19. Zhao, J., J. Wu, X. Feng, H. Xiong, and K. Xu. Information propagation in online social networks: a tie-strength perspective. *Knowledge and Information Systems*, Vol. 32, No. 3, Nov. 2011, pp. 589–608.

20. Sosa, M. E. Where Do Creative Interactions Come From? The Role of Tie Content and Social Networks. *Organization Science*, Vol. 22, No. 1, Feb. 2011, pp. 1–21.

21. Granovetter, M. S. The Strength of Weak Ties. *American Journal of Sociology*, Vol. 78, No. 6, 1973, pp. 1360–1380.

22. Retzer, S., P. Yoong, and V. Hooper. Inter-organisational knowledge transfer in social networks: A definition of intermediate ties. *Information Systems Frontiers*, Vol. 14, No. 2, Jul. 2010, pp. 343–361.

23. Arnaboldi, V., A. Guazzini, and A. Passarella. Egocentric online social networks: Analysis of key features and prediction of tie strength in Facebook. *Computer Communications*, Vol. 36, No. 10-11, Jun. 2013, pp. 1130–1144.

24. Gilbert, E., and K. Karahalios. Predicting tie strength with social media. *Proceedings of the 27th international conference on Human factors in computing systems - CHI 09*, 2009, p. 211.

25. Pappalardo, L., G. Rossetti, and D. Pedreschi. “How Well Do We Know Each Other?” Detecting Tie Strength in Multidimensional Social Networks. *2012 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, Aug. 2012, pp. 1040–1045.

26. He, Y., C. Zhang, and Y. Ji. Principle Features for Tie Strength Estimation in Micro-blog Social Network. *2012 IEEE 12th International Conference on Computer and Information Technology*, Oct. 2012, pp. 359–367.

27. Marsden, P., P. Carrington, J. Scott, and S. Wasserman. Recent developments in network measurement. In *Models and methods in social network analysis*, Cambridge University Press, New York, pp. 8–30.

28. Degenne, A., and M. Forse. *Introducing social networks*. Sage, London, 1999.

29. Berg, P., T. Arentze, and H. Timmermans. A multilevel path analysis of contact frequency between social network members. *Journal of Geographical Systems*, Vol. 14, No. 2, Oct. 2010, pp. 125–141.

30. Axhausen, K. W. Social networks, mobility biographies, and travel: survey challenges. *Environment and Planning B: Planning and Design*, Vol. 35, No. 6, 2008, pp. 981–996.

31. Wellman, B., and S. Wortley. Different Strokes from Different Folks : Community Ties and Social Support. *American Journal of Sociology*, Vol. 96, No. 3, 1990, pp. 558–588.

32. Bramoullé, Y., H. Djebbari, and B. Fortin. Identification of peer effects through social networks. *Journal of Econometrics*, Vol. 150, No. 1, May 2009, pp. 41–55.

33. Calvó-Armengol, A., E. Patacchini, and Y. Zenou. Peer effects and social networks in education. *Review of Economic Studies*, Vol. 76, No. 4, Oct. 2009, pp. 1239–1267.

34. Adjemian, M. K., C.-Y. C. Lin, and J. Williams. Estimating spatial interdependence in automobile type choice with survey data. *Transportation Research Part A: Policy and Practice*, Vol. 44, No. 9, Nov. 2010, pp. 661–675.

35. Bhat, C. R., R. Paleti, and P. Singh. A Spatial Multivariate Count Model for Firm Location Decisions. *Journal of Regional Science*, Vol. In press, 2013.

36. R Core Team. R Project.

37. ESRI. ESRI. www.esri.com.

38. LeSage, J. P., and R. K. Pace. *Introduction to Spatial Econometrics*. Chapman and Hall/CRC, 2009.

39. Adamic, L. a, and E. Adar. Friends and neighbors on the Web. *Social Networks*, Vol. 25, No. 3, Jul. 2003, pp. 211–230.

40. Feld, S. L. Why Your Friends Have More Friends than You Do. *American Journal of Sociology*, Vol. 96, No. 6, 1991, pp. 1464–1477.

41. Paleti, R., C. R. Bhat, R. M. Pendyala, and K. G. Goulias. The Modeling of Household Vehicle Type Choice Accommodating Spatial Dependence Effects.