Quantifying the Attractiveness of Destinations for Modeling Long Distance Travel Patterns

Rui Zhu

STKO Lab
Department of Geography
University of California, Santa Barbara

Outline

I. Long Distance Travel Patterns

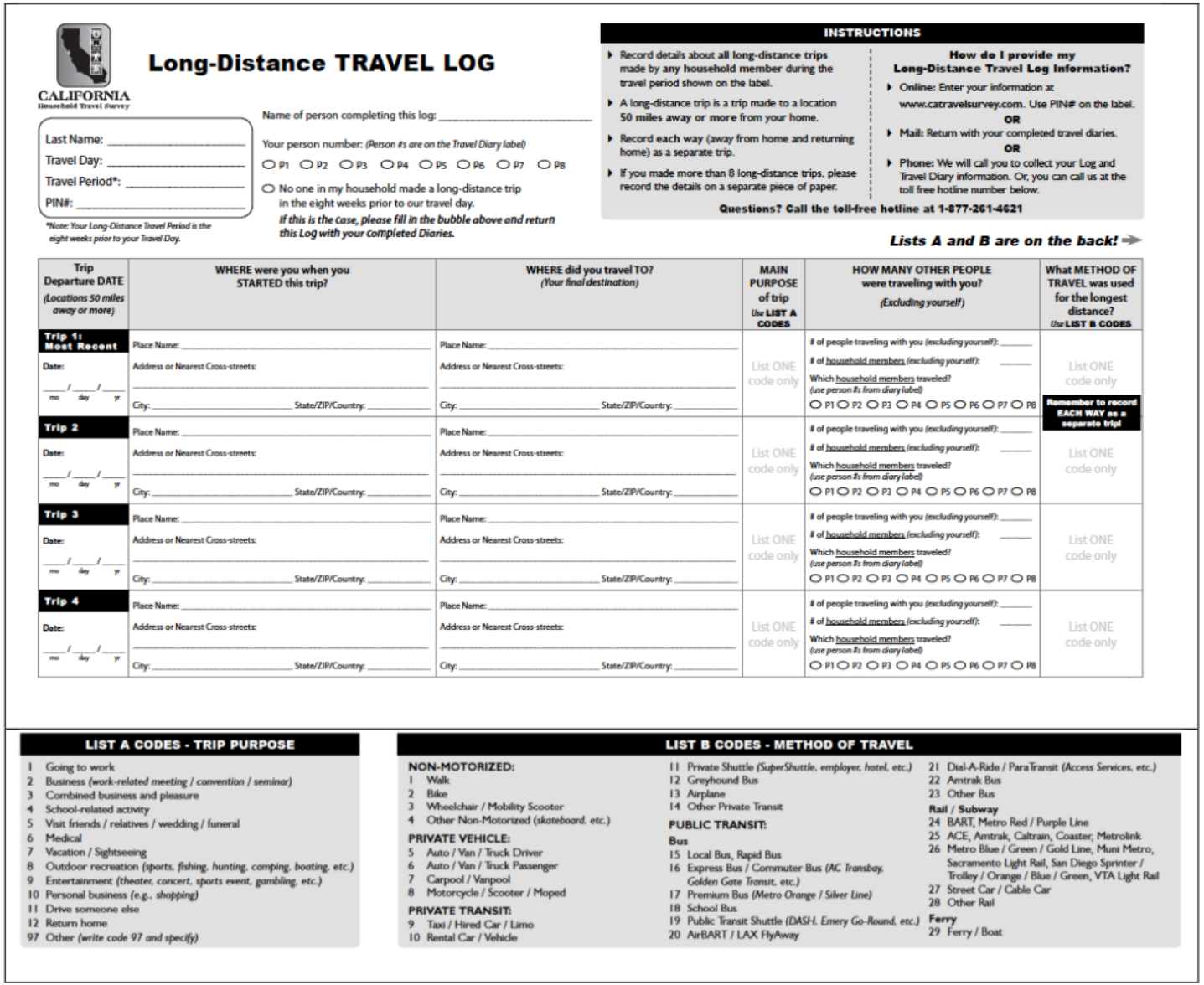
II. Semantic Signatures

I. Long Distance Travel Patterns

– With destination attractiveness as one geographic context

Motivations

- California Household Travel Survey (CHTS)
- develop **profiles of long distance travel in California** at a household scale to be used in a variety of planning activities



CALIFORNIA
Household Travel Survey

Long-Distance TRAVEL LOG

Last Name: _____
Travel Day: _____
Travel Period*: _____
PIN#: _____

*Note: Your Long-Distance Travel Period is the eight weeks prior to your Travel Day.

Name of person completing this log: _____

Your person number: *(Person #s are on the Travel Diary label)*

○ P1 ○ P2 ○ P3 ○ P4 ○ P5 ○ P6 ○ P7 ○ P8

○ No one in my household made a long-distance trip in the eight weeks prior to our travel day.
If this is the case, please fill in the bubble above and return this Log with your completed Diaries.

INSTRUCTIONS

- Record details about **all** long-distance trips made by any household member during the travel period shown on the label.
- A long-distance trip is a trip made to a location **50 miles away or more** from your home.
- Record **each way** (away from home and returning home) as a separate trip.
- If you made more than 8 long-distance trips, please record the details on a separate piece of paper.

How do I provide my Long-Distance Travel Log Information?

- **Online:** Enter your information at www.catravelsurvey.com. Use PIN# on the label.
  OR
- **Mail:** Return with your completed travel diaries.
  OR
- **Phone:** We will call you to collect your Log and Travel Diary information. Or, you can call us at the toll free hotline number below.

Questions? Call the toll-free hotline at 1-877-261-4621

Lists A and B are on the back!

Trip Departure DATE *(Locations 50 miles away or more)*	WHERE were you when you STARTED this trip?	WHERE did you travel TO? *(Your final destination)*	MAIN PURPOSE of trip *Use* **LIST A CODES**	HOW MANY OTHER PEOPLE were traveling with you? *(Excluding yourself)*	What METHOD OF TRAVEL was used for the longest distance? *Use* **LIST B CODES**
Trip 1: Most Recent Date: ___/___/___ mo day yr	Place Name: _____ Address or Nearest Cross-streets: _____ City: _____ State/ZIP/Country: _____	Place Name: _____ Address or Nearest Cross-streets: _____ City: _____ State/ZIP/Country: _____	List ONE code only	# of people traveling with you *(excluding yourself)*: _____ # of household members *(excluding yourself)*: _____ Which household members traveled? *(use person #s from diary label)* ○ P1 ○ P2 ○ P3 ○ P4 ○ P5 ○ P6 ○ P7 ○ P8	List ONE code only **Remember to record EACH WAY as a separate trip!**
Trip 2 Date: ___/___/___ mo day yr	Place Name: _____ Address or Nearest Cross-streets: _____ City: _____ State/ZIP/Country: _____	Place Name: _____ Address or Nearest Cross-streets: _____ City: _____ State/ZIP/Country: _____	List ONE code only	# of people traveling with you *(excluding yourself)*: _____ # of household members *(excluding yourself)*: _____ Which household members traveled? *(use person #s from diary label)* ○ P1 ○ P2 ○ P3 ○ P4 ○ P5 ○ P6 ○ P7 ○ P8	List ONE code only
Trip 3 Date: ___/___/___ mo day yr	Place Name: _____ Address or Nearest Cross-streets: _____ City: _____ State/ZIP/Country: _____	Place Name: _____ Address or Nearest Cross-streets: _____ City: _____ State/ZIP/Country: _____	List ONE code only	# of people traveling with you *(excluding yourself)*: _____ # of household members *(excluding yourself)*: _____ Which household members traveled? *(use person #s from diary label)* ○ P1 ○ P2 ○ P3 ○ P4 ○ P5 ○ P6 ○ P7 ○ P8	List ONE code only
Trip 4 Date: ___/___/___ mo day yr	Place Name: _____ Address or Nearest Cross-streets: _____ City: _____ State/ZIP/Country: _____	Place Name: _____ Address or Nearest Cross-streets: _____ City: _____ State/ZIP/Country: _____	List ONE code only	# of people traveling with you *(excluding yourself)*: _____ # of household members *(excluding yourself)*: _____ Which household members traveled? *(use person #s from diary label)* ○ P1 ○ P2 ○ P3 ○ P4 ○ P5 ○ P6 ○ P7 ○ P8	List ONE code only

LIST A CODES - TRIP PURPOSE

1 Going to work
2 Business *(work-related meeting / convention / seminar)*
3 Combined business and pleasure
4 School-related activity
5 Visit friends / relatives / wedding / funeral
6 Medical
7 Vacation / Sightseeing
8 Outdoor recreation *(sports, fishing, hunting, camping, boating, etc.)*
9 Entertainment *(theater, concert, sports event, gambling, etc.)*
10 Personal business *(e.g., shopping)*
11 Drive someone else
12 Return home
97 Other *(write code 97 and specify)*

LIST B CODES - METHOD OF TRAVEL

NON-MOTORIZED:
1 Walk
2 Bike
3 Wheelchair / Mobility Scooter
4 Other Non-Motorized *(skateboard, etc.)*

PRIVATE VEHICLE:
5 Auto / Van / Truck Driver
6 Auto / Van / Truck Passenger
7 Carpool / Vanpool
8 Motorcycle / Scooter / Moped

PRIVATE TRANSIT:
9 Taxi / Hired Car / Limo
10 Rental Car / Vehicle
11 Private Shuttle *(SuperShuttle, employer, hotel, etc.)*
12 Greyhound Bus
13 Airplane
14 Other Private Transit

PUBLIC TRANSIT:
Bus
15 Local Bus, Rapid Bus
16 Express Bus / Commuter Bus *(AC Transbay, Golden Gate Transit, etc.)*
17 Premium Bus *(Metro Orange / Silver Line)*
18 School Bus
19 Public Transit Shuttle *(DASH, Emery Go-Round, etc.)*
20 AirBART / LAX FlyAway
21 Dial-A-Ride / ParaTransit *(Access Services, etc.)*
22 Amtrak Bus
23 Other Bus
Rail / Subway
24 BART, Metro Red / Purple Line
25 ACE, Amtrak, Caltrain, Coaster, Metrolink
26 Metro Blue / Green / Gold Line, Muni Metro, Sacramento Light Rail, San Diego Sprinter / Trolley / Orange / Blue / Green, VTA Light Rail
27 Street Car / Cable Car
28 Other Rail
Ferry
29 Ferry / Boat

Data - Survey

- 18,008 (42.44%) from the total of 42,431 CHTS households

	HH Size	# of Workers	# of Students	# of Drivers	# of Cars	# of Daily Trips in Diary	Total HHs in CHTS
Households that Report Long Distance Travel	2.59	1.31	0.65	1.98	2.02	9.36	18,008
Households that do not Report* Long Distance Travel	2.56	1.15	0.64	1.78	1.75	7.50	24,423

Preliminary Analysis

1. Binary Logit model → model whether a household has LD travel or not using its basic characteristics

Selected results: Hispanic households are more likely to report LD travel but when the interview was done in Spanish they are less likely to report LD travel.

1. Zero-inflated Poisson/Negative Binomial regression→ model the count of LD

Selected results: the more persons/workers in a household, the fewer number of LD travels → intra-household constraints.

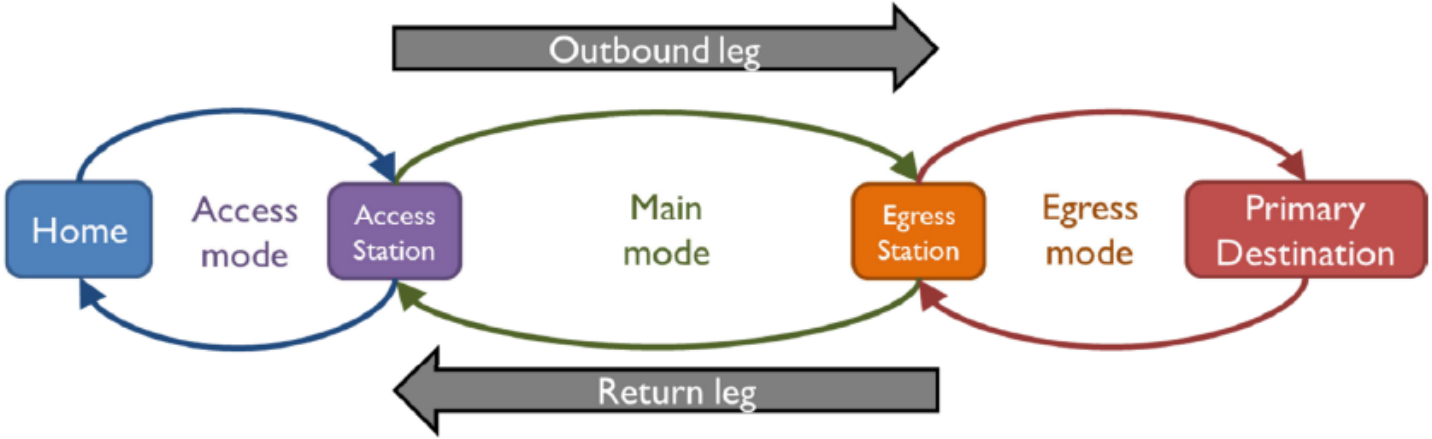
Contexts (household)

(dependent variable: 0 = reported long distance and 1 did not report long distance trips)

	Estimate	Std. Error	z value
(Intercept)	1.256	0.077	16.324
Household Size	0.078	0.015	5.347
# Workers	-0.034	0.014	-2.397
# Students	-0.096	0.017	-5.676
# Cars	-0.101	0.013	-7.650
Interview in Spanish	0.610	0.063	9.749
Hispanic Household	-0.198	0.028	-7.190
Household Income is			
$10,000 to $24,999	-0.168	0.070	-2.389
$25,000 to $34,999	-0.390	0.073	-5.365
$35,000 to $49,999	-0.534	0.070	-7.585
$50,000 to $74,999	-0.736	0.069	-10.695
$75,000 to $99,999	-0.921	0.070	-13.100
$100,000 to $149,999	-1.055	0.071	-14.967
$150,000 to $199,999	-1.076	0.077	-13.983
$200,000 to $249,999	-1.179	0.089	-13.225
$250,000 or more	-1.199	0.088	-13.662
Income Don't Know	-0.223	0.091	-2.444
Income Refused	-0.702	0.076	-9.180
Does not own home	0.141	0.033	4.326
Lives in single family home	-0.082	0.033	-2.462
Lives in mobile home	0.190	0.070	2.723
Lives in bldg with 20+ apartments	0.210	0.053	3.984

How about external (geographic) contexts ?

- However, CHTS provides data about the main outbound leg and return leg but not the access and egress portion of a tour; it also does not provide information about opportunities for **activity participation** at the access and egress station, home location and primary destination.



Long Distance Conceptual Tour Structure (reproduced from CSI, 2014)

Geographic Contexts (Semantic Signatures)

- Why is this destination **attractive** to the household?

  Data: Foursquare
  Method: the N nearest POIs are queried for each destination

  - Popularity of the neighborhood
    - check-ins, number of users, number of tips, ratings (quantiles, distance-weighted statistics, distance-buffering statistics)
  - Semantics of the neighborhood
    - feature types
    - convex area of the nearest POIs → spatial arrangement

Geographic Contexts (Semantic Signatures)

- Feature types → topics of the neighborhood

L. A. INTERNATIONAL AIRPORT, LOS ANGELES, CA, -118.402943, 33.942878

HOME, SANTA BARBARA, CA, -119.805985, 34.452963

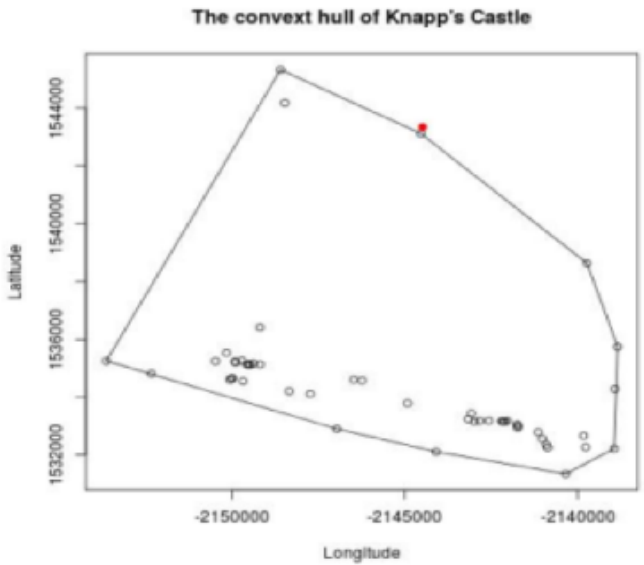
BIKE RACE, REDDING, CA, -122.359935, 40.569064

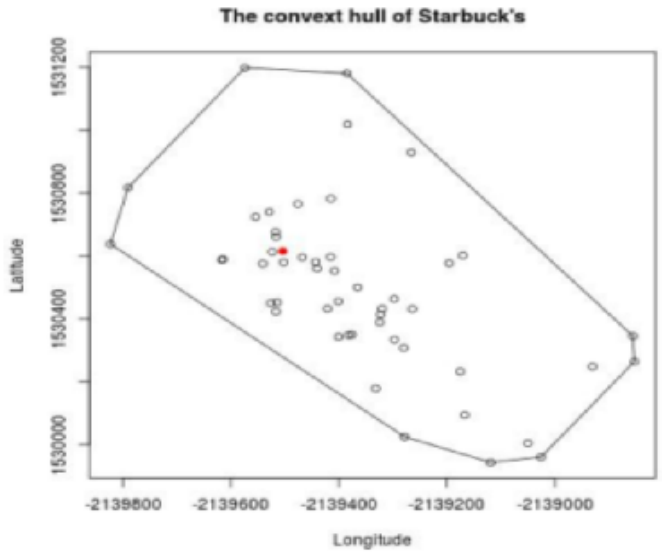
HAPUNA BEACH PRINCE HOTEL, KAMUELA, HI, -155.821209, 20.002854

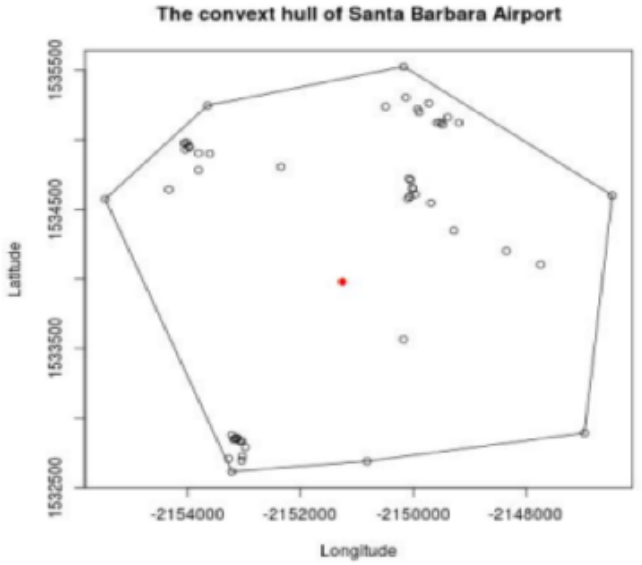
- **Entropy** of feature type: the diversity of the neighborhood
- **Topic modeling (Latent Dirichlet Allocation)**: as an analogue to the text mining, we regarded each destination as an document, with its N neighbors' feature types as the words in the document → 50 topic distribution for each destination

Geographic Contexts (Semantic Signatures)

- Convex hull:







Modeling the LD pattern

- Behavior outcome variables (dependents)
  - **Transportation Mode**: miles traveled by air, by car, by public transportation, Overnight tour: overnight stays of the tour;
  - **Temporal**: the season during which the tour was made,
  - **Spatial**: indicator whether the tour was in CA
  - **Purposes**: the main purpose of the tour, number of trips in work related business, vacation, and outdoor recreation.
- Behavior determinate variables (independents)
  - **Household**: home type, household size, number of employed, number of vehicles, home type, Hispanic, block group category …
  - **Foursquare**: rating, number of users, convex hull area ...

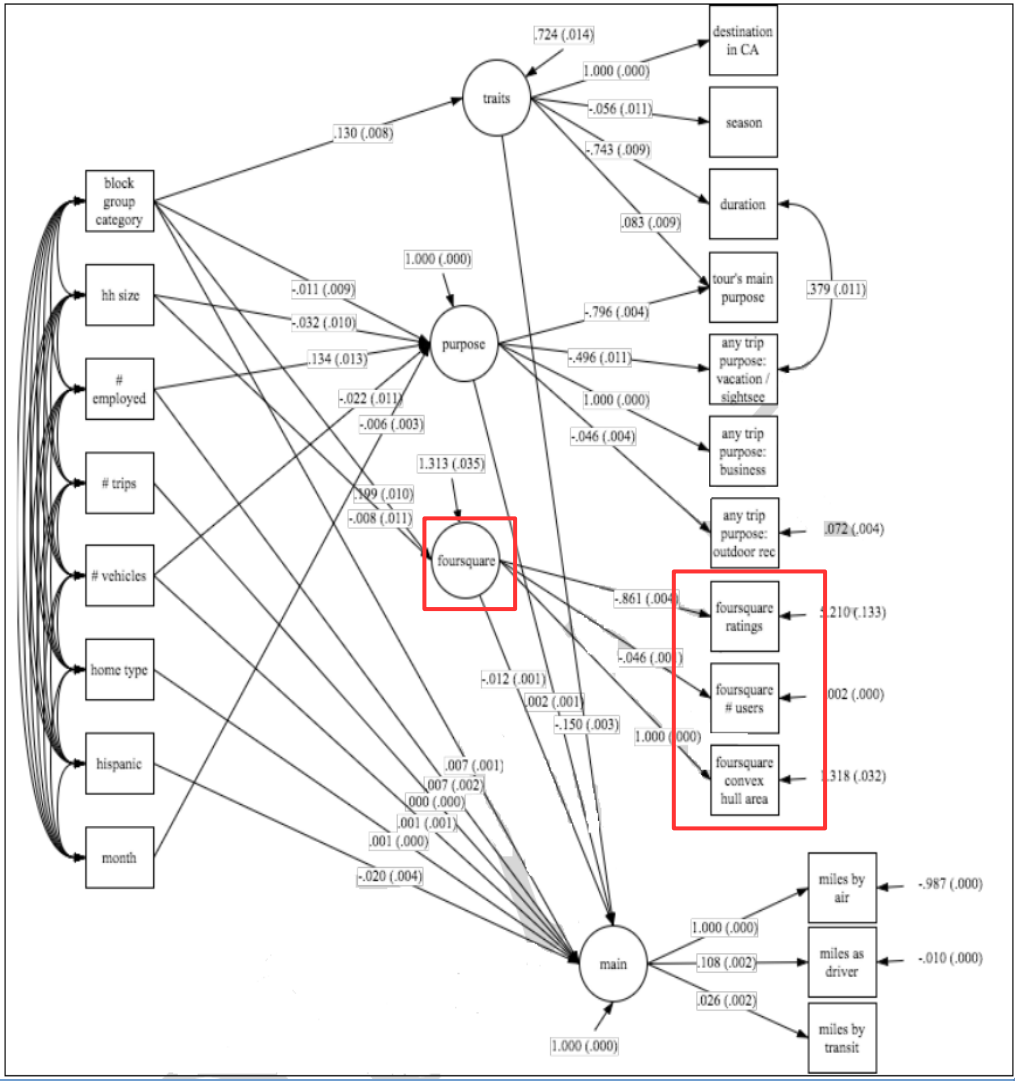
Modeling the LD pattern

- Structural Equation Model (with latent variables)

Measurement model for y: $y = \Lambda_y \eta + \varepsilon$

Measurement model for x: $x = \Lambda_x \xi + \delta$

Structural model: $\eta = B\eta + \Gamma\xi + \zeta$



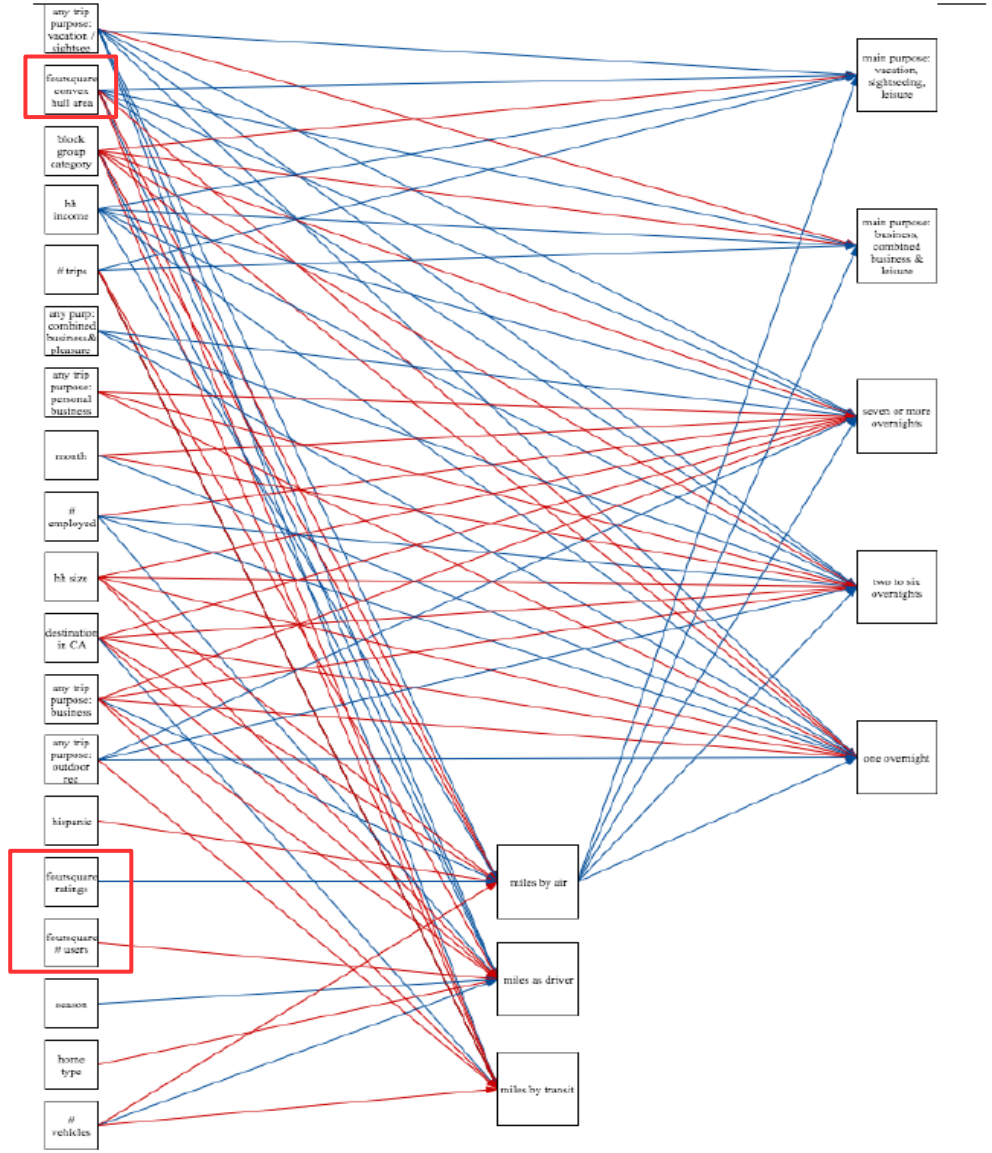
Observations

- Overall we see that in addition to the household characteristics influencing long distance travel, the **place and type of residence** play an important and significant role in shaping long distance travel patterns.
- We also see that **foursquare** does provide significant indicators of attractiveness of the main travel destination and shows it is worthwhile continuing the collection of data of this type via an observatory that combines behavioral data from diaries with land use data of the residence and social media data.

Modeling the LD pattern

- Structure Equation Model (without latent variables)

$$y = By + \Gamma x + \zeta$$



Observations

- **Higher income** are more likely to **travel by air** for longer distances but larger **households with many cars**, living in **exurbs and rural environments**, and **Hispanic** households are less likely to travel by air long distances. They are more likely to fly long distance when one or more trips in a tour are for work related **business and/or vacation**.
- The foursquare relationships show tours with more air miles are more likely to be in **denser areas** (e.g., big cities) that received **higher foursquare ratings**.

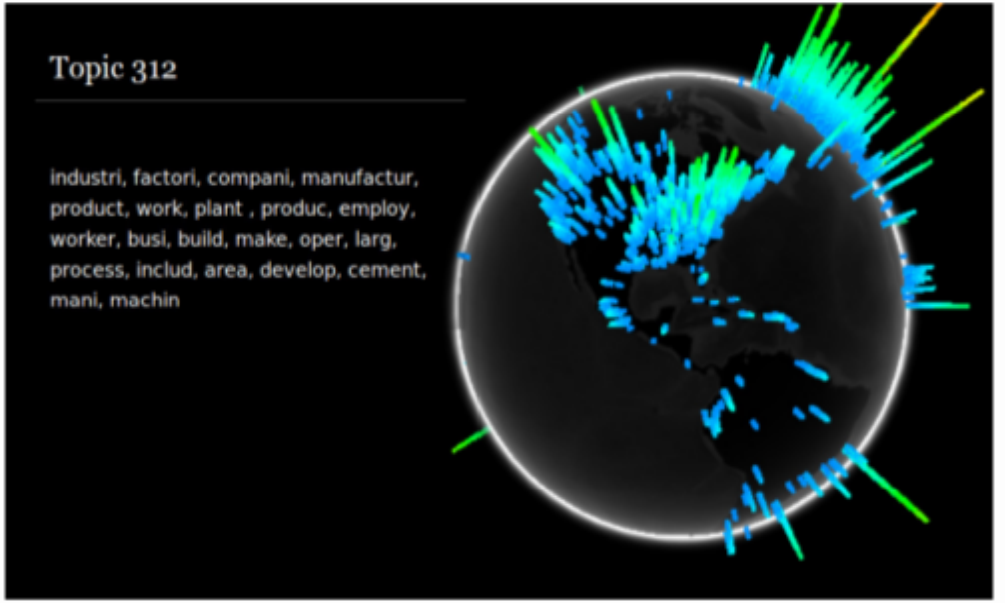
II. GeoSpatial Semantic Signatures

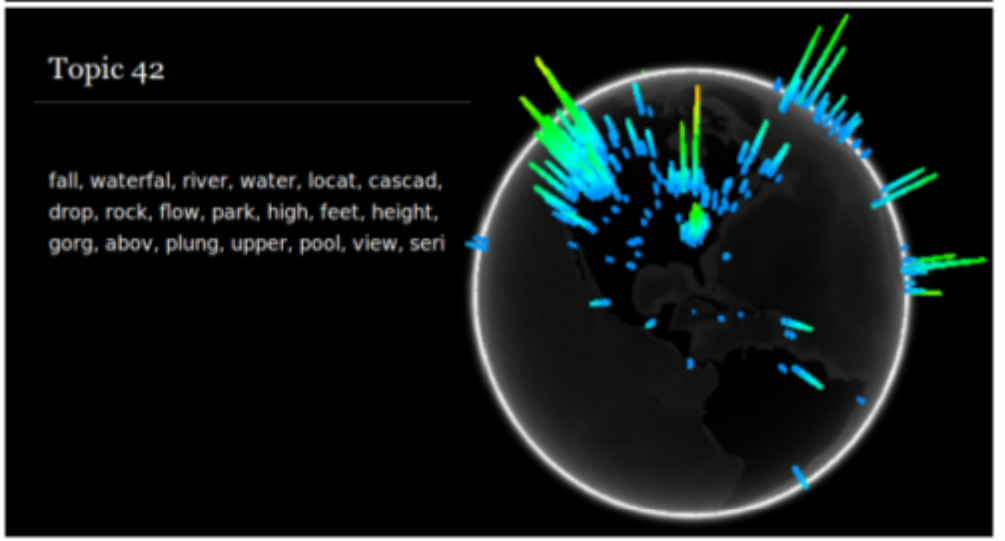
- A framework to quantify geographic contexts

Semantic Signatures

- **Thematic** Signature:

  can be computed out of unstructured text using **topic modelings** from sources such as Wikipedia, travel blogs, news article, and so forth



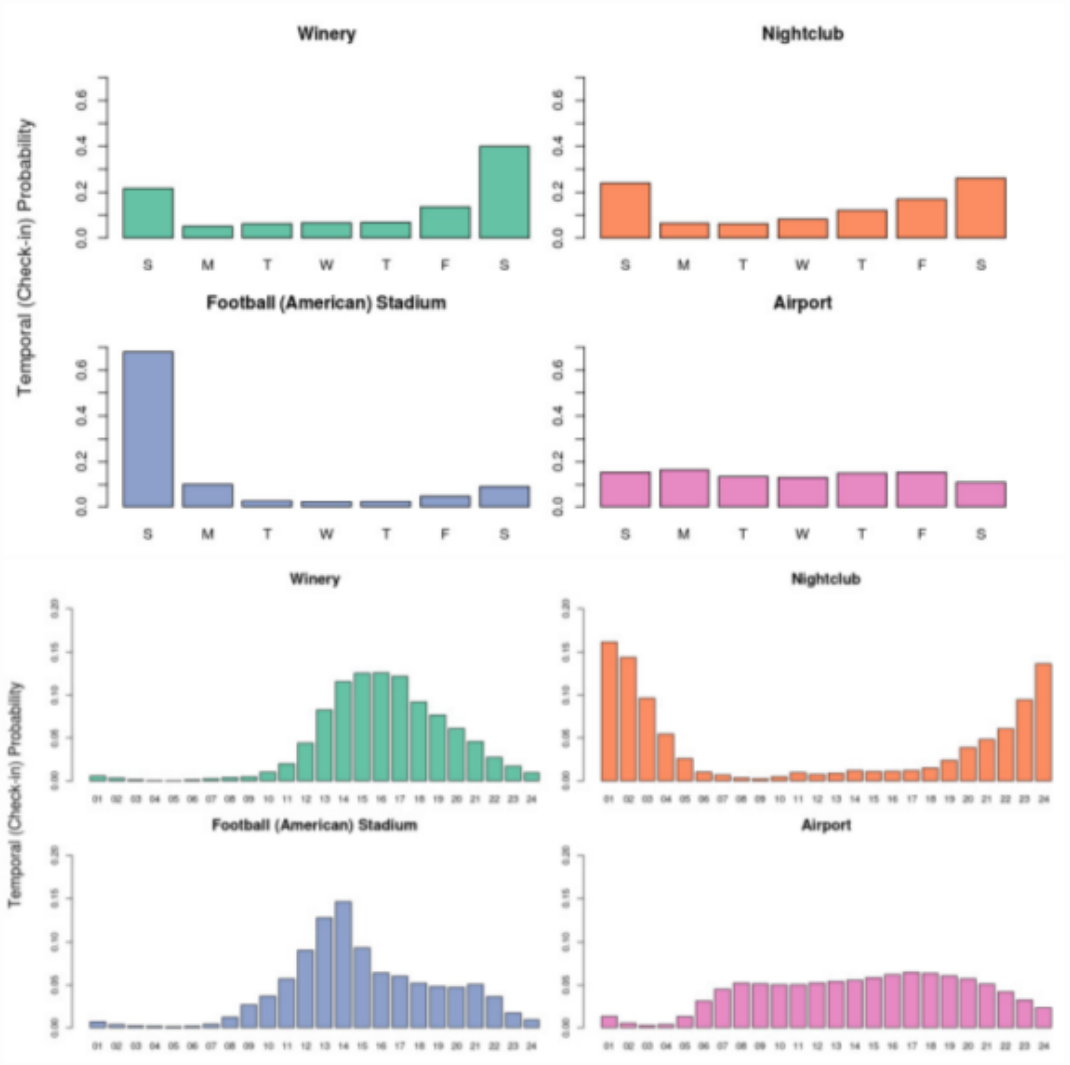


Adams, B. and Janowicz, K. (2015) Thematic Signatures for Cleansing and Enriching Place-Related Linked Data. International Journal of Geographical Information Science. Volume 29, Issue 4.

Semantic Signatures

- **Temporal** Signatures:

  analyze on the **time trend** of checkins, or ant other activities, at an instance/feature type using social media data



McKenzie, G., and Janowicz, K. (2015) Where is also about time: A location-distortion model to improve reverse geocoding using behavior-driven temporal signatures. Computers, Environment and Urban Systems. Volume 54, November 2015, pp. 1-13

Semantic Signatures

- Spatial Signatures:

  use **spatial statistics** to quantify the spatial organization of places

Spatial Point Pattern		Spatial Association	Spatial Interaction with Other Geographic Features		Place-based Statistics	
Local	Intensity	Global Moran' I	Population	min	Number of distinct states (or counties)	
	Mean distance to nearest neighbor			max		
	Standard deviation of distance to nearest neighbor			mean	Entropy of states (or counties)	
	Kernel density (range)			std.		
		Semivariogram value (at first distance lag)	Road Network	min of shortest distance	Number of adjacency states (or counties) that also have features with the target type	
	Kernel density (bandwidth)			max of shortest distance		
	Ripley's K (range)			mean of shortest distance	Number of distinct feature types for nearest neighbor	
	Ripley's K (mean deviation)			std. of shortest distance		
		Semivariogram value (at median distance lag)		Entropy of nearest road types	Entropy of feature types for nearest neighbor	
	Standard deviation ellipse (rotation)		Climate	Mean precipitation		
	Standard deviation ellipse (std. along x-axis)			Std. precipitation	LDA-based Approach	Mean KL Divergence of the topic distribution
	Standard deviation ellipse (std. along y-axis)			Mean Tmax		Entropy of the topic distribution
		Semivariogram value (at last distance lag)		Std. Tmax		
Global	Intensity			Mean Tmin		
	Kernel density (range)			Mean Tmax		
	Kernel density (band width)			Mean Water vapor pressure		
				Std. Water vapor pressure		

Zhu, R., Hu, Y., Janowicz, K. , and McKenzie, G. (2016) Spatial Signatures for Geographic Feature Types: Examining Gazetteer Ontologies using Spatial Statistics. Transactions in GIS

Applications

- Current applications of using signatures to integrate geographic contexts
  - Place name disambiguation
  - Map user locations from spaces to places
  - Geo-spatial ontology / feature type schema alignment
  - Coreference resolution
  - **Place attractiveness for modeling travel patterns**

Conclusions & Discussions

- Discussed our initial results of modeling long distance travel patterns at California with context augmentation at destinations;
- Introduced semantic signatures for quantifying geographic contexts.

However, there are still some open questions:

- How to use semantic signatures, or the guideline we provided, to quantify contexts for modeling mobility (for both human and animals)?
- How to model the trajectory context?

Acknowledgements

University of California Center on Economic Competitiveness in Transportation

Geotrans
The Transportation Portal @ UC Santa Barbara

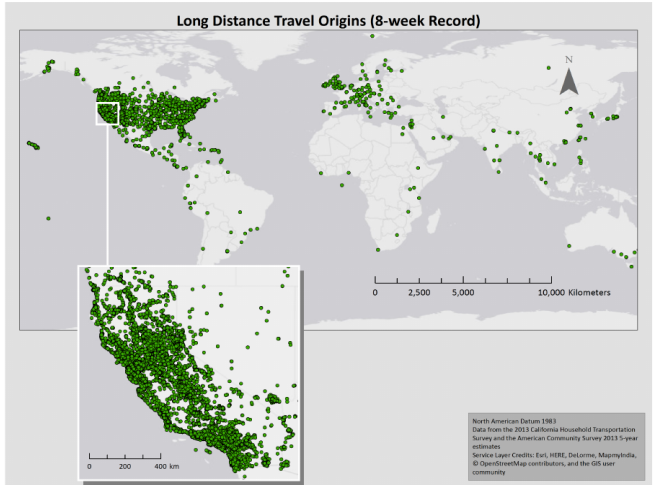
Collaborators: Krzysztof Janowicz, Konstadinos Goulias, Yingjie Hu, Grant McKenzie, Song Gao, Bao Yan, Yiting Ju, Gengchen Mai, Adam W. Davis, Elizabeth C. McBride

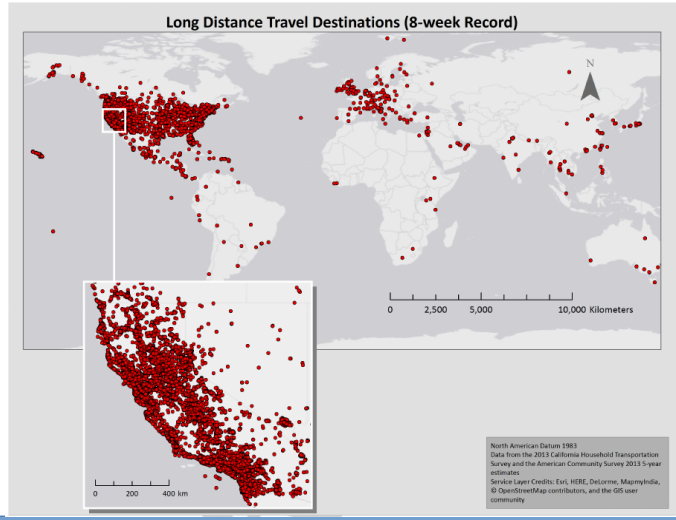
Thanks for your attention!

Question/comments?

Tour Completeness Category	Criteria	Potentially Usable Characteristics
Single Trip	• One trip • No continuity with other LD trips reported by this household	• Origin and destination locations and distance from home • Single mode used • Single trip purpose
Partial Tour	• Multiple trips in sequence • Trip destinations generally match next origin, but discontinuity is allowed • Start and/or end may not be at home	• Duration (floor only) • Multiple locations • Mix of modes • Mix of purposes
Complete Tour	• Multiple trips in sequence • Trip destinations always match next origin • Origin of first trip and destination of last trip are at home	• Duration (precise) • Multiple locations • Mix of modes • Mix of purposes

	Single Trip	Partial Tour	Complete
Single Day	--	970	7,688
One Overnight	--	593	1,924
2-6 Overnights	--	1,545	5,207
7+ Overnights	--	1,979	2,122
Unknown Duration	18,298	172	678
Total Tours	18,298	5,259	17,619
Total Trips	18,298	12,690	37,194
Households with at least one such trip	9,513	3,560	9,540





Measurement model for y: $y = \Lambda_y \eta + \varepsilon$ (Eq I.1)

Measurement model for x: $x = \Lambda_x \xi + \delta$ (Eq I.2)

Structural model: $\eta = \mathrm{B}\eta + \Gamma\xi + \zeta$ (Eq. I.3)

where $y = p \times 1$ vector of observed endogenous variables.

$x = q \times 1$ vector of observed exogenous variables.

$\eta = m \times 1$ vector of latent endogenous variables.

$\xi = n \times 1$ vector of latent exogenous variables.

$\varepsilon = p \times 1$ vector of measurement errors in y.

$\delta = q \times 1$ vector of measurement errors in x.

$\Lambda_y = p \times m$ matrix of coefficients of the regression of y on η.

$\Lambda_x = q \times n$ matrix of coefficients of the regression of x on ξ.

$\mathrm{B} = m \times m$ matrix of coefficients of the η-variables in the structural relationships.

$\Gamma = m \times n$ matrix of coefficients of the ξ-variables in the structural relationships.

$\zeta = m \times 1$ vector of equation errors in the structural relationships.

$y = \mathrm{B}y + \Gamma x + \zeta$ (Eq. I.4)

where $y = p \times 1$ vector of observed endogenous variables.

$x = q \times 1$ vector of observed exogenous variables.

$\mathrm{B} = p \times p$ matrix of coefficients of the y-variables.

$\Gamma = p \times q$ matrix of coefficients of the x-variables.

$\zeta = p \times 1$ vector of equation errors.