



Quantifying Semantic Uncertainties of Volunteered Geographic Information to Understand Human Travel Behaviors

Rui Zhu, Krzysztof Janowicz

STKO Lab

University of California, Santa Barbara



Outline

- Motivation
- Methodology
- Experiment
- Discussion


Motivations

Conventional Data Source:

Internal resource →

Household Travel Survey

- Income
- Household size
- # of daily trips
-



Long-Distance TRAVEL LOG

Name of person completing this log: _____

Last Name: _____

Travel Day: _____

Travel Period*: _____

PIN#: _____

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

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	Place Name: _____ Date: ____/____/____ 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 (including yourself): _____ # of household members (including yourself): _____ Which household members traveled? (see person #s from diary label) <input type="radio"/> P1 <input type="radio"/> P2 <input type="radio"/> P3 <input type="radio"/> P4 <input type="radio"/> P5 <input type="radio"/> P6 <input type="radio"/> P7 <input type="radio"/> P8	List ONE code only Remember to record EACH WAY as a separate trip!
Trip 2	Place Name: _____ Date: ____/____/____ 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 (including yourself): _____ # of household members (including yourself): _____ Which household members traveled? (see person #s from diary label) <input type="radio"/> P1 <input type="radio"/> P2 <input type="radio"/> P3 <input type="radio"/> P4 <input type="radio"/> P5 <input type="radio"/> P6 <input type="radio"/> P7 <input type="radio"/> P8	List ONE code only
Trip 3	Place Name: _____ Date: ____/____/____ 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 (including yourself): _____ # of household members (including yourself): _____ Which household members traveled? (see person #s from diary label) <input type="radio"/> P1 <input type="radio"/> P2 <input type="radio"/> P3 <input type="radio"/> P4 <input type="radio"/> P5 <input type="radio"/> P6 <input type="radio"/> P7 <input type="radio"/> P8	List ONE code only
Trip 4	Place Name: _____ Date: ____/____/____ 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 (including yourself): _____ # of household members (including yourself): _____ Which household members traveled? (see person #s from diary label) <input type="radio"/> P1 <input type="radio"/> P2 <input type="radio"/> P3 <input type="radio"/> P4 <input type="radio"/> P5 <input type="radio"/> P6 <input type="radio"/> P7 <input type="radio"/> 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:

 - 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 Transit, 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

PRIVATE VEHICLE:

 - 5 Auto / Van / Truck Driver
 - 6 Auto / Van / Truck Passenger
 - 7 Carpool / Vanpool
 - 8 Motorcycle / Scooter / Moped

PRIVATE TRAVEL:

 - 9 Taxi / Hired Car / Limo
 - 10 Rental Car / Vehicle

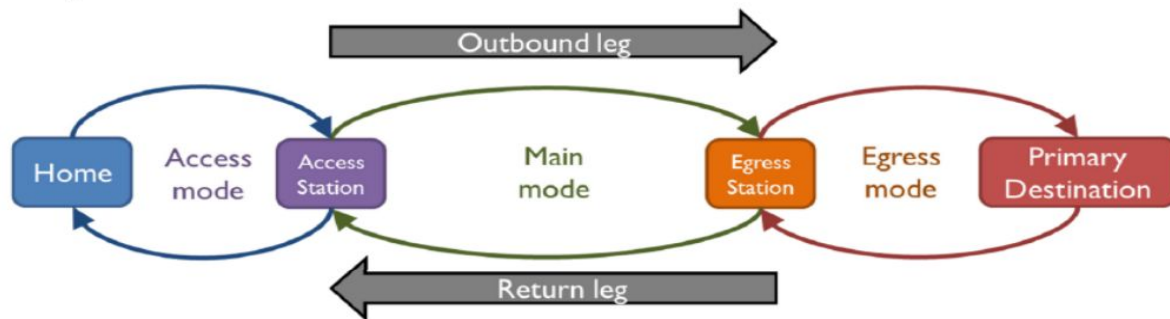
Other:

 - 21 Dial-A-Ride / ParaTransit (Access Services, etc.)
 - 22 Amtrak Bus
 - 23 Other Bus
 - 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
 - 29 Ferry / Boat

Motivations

However:

- Internal resource (e.g., household travel survey) 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 stations, home location and primary destination.





Motivations

How about external contexts of travels?

- Understanding of the place/neighborhood (originations, destinations, or intermediate stops)
- Data source: Volunteered Geographic Information (VGI)



Motivations

- Check-in numbers; number of reviews; ratings → Popularity
- **Place Types** → **Functionality/Human's perception**

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



HOME, SANTA BARBARA, CA, -119.805985, 34.452963



How to quantify semantics?

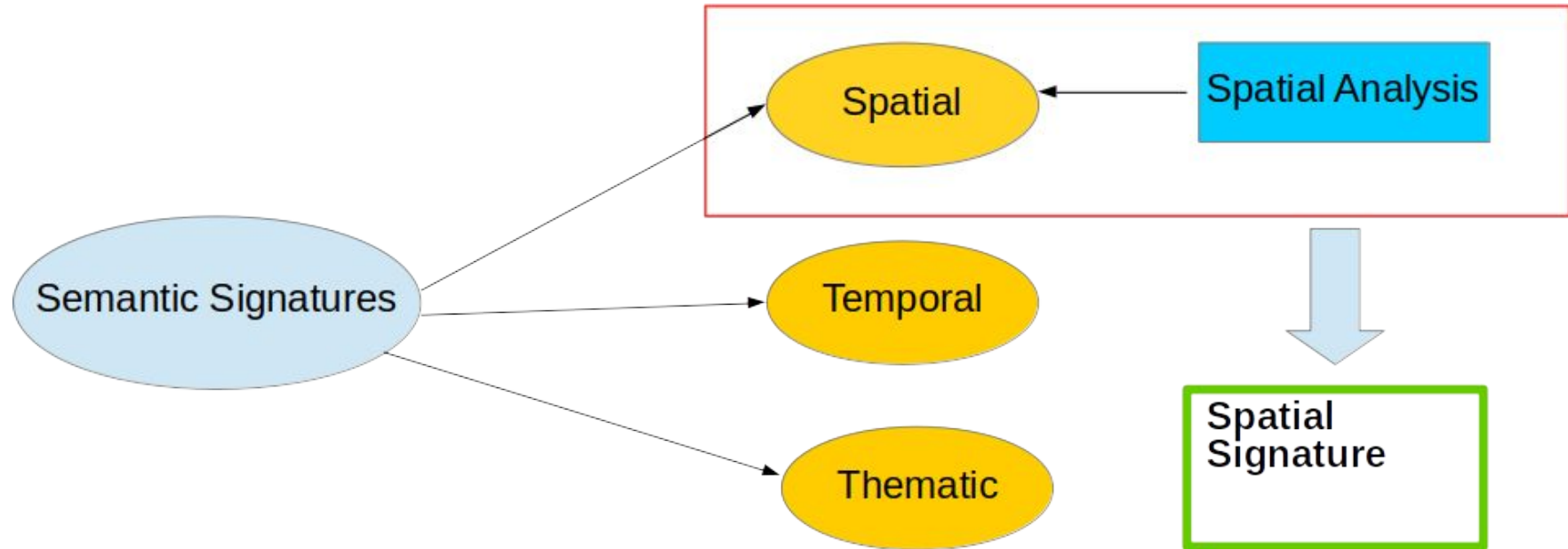
BIKE RACE, REDDING, CA, -122.359935, 40.569054



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



Methodology



Janowicz, K., McKenzie, G., Hu, Y., Zhu, R., and Gao, So. (2018): [Using Semantic Signatures for Social Sensing in Urban Environments](#). Mobility Patterns, Big Data and Transport Analytics.



Methodology

Spatial Signatures

- ***Spatial structure*** of the data belonging to a place type is used to quantify its semantics.
- ***Spatial statistics*** are applied to describe such spatial structure.
- Spatial point patterns. Spatial autocorrelation analysis, spatial interaction analysis with other geographic features, place-based analysis. → **41 statistics**

Methodology

Spatial Signature -- Spatial point patterns

- Intensity-based: local intensity, kernel density estimation
- Distance-based: nearest-neighbor distance, Ripley's K, and standard deviational analysis

Randomly Selected Points using CSR in Contiguous US



Generate random points
(Complete Spatial Randomness)

Geonames Dens in Contiguous US



Spatial Point Pattern for the 400 Randomly Selected Sample
(Geonames Dens)



Select nearest 100 neighbors
for each random points



Kernel Density Estimation for the 400 Randomly Selected Sample



Average the statistics
over all random points

Standard Deviation Ellipse of Spatial Point Pattern
for the 400 Randomly Selected Sample (Geonames Dens)



Conduct spatial point pattern
analysis on these 100 neighbors

Methodology

Spatial Signature -- Spatial point patterns -- Examples

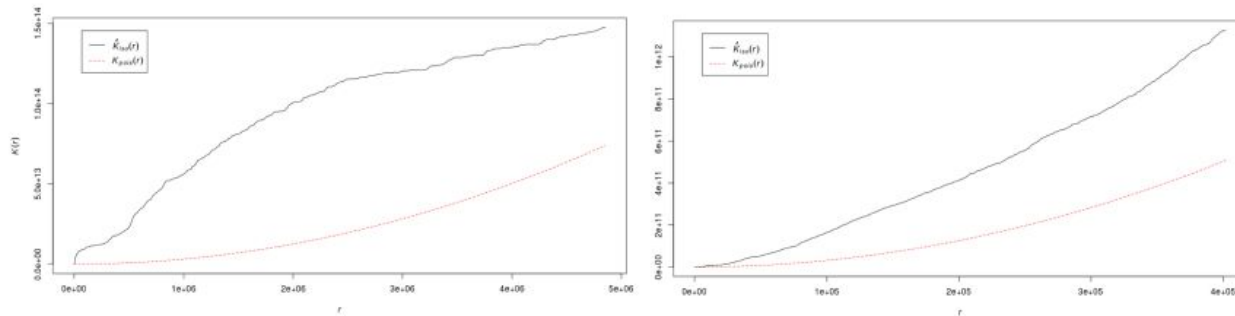


Figure 1: Ripley's K of *Park* (left) and *Dam* (right) from DBpedia Places.

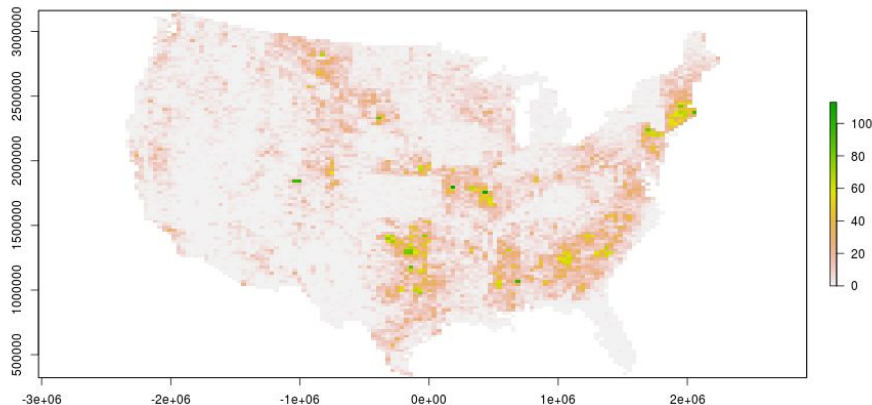
Statistics: mean and std. of the deviation between theoretical can observed K curves

Methodology

Spatial Signature -- Spatial AutoCorrelation Analysis

- Moran's I: how intensities of cells differ from their neighbors
- Semivariogram: measure the variation of cell intensities in a specific distance lag class.

Dams in GeoNames



Cell size : 36 km * 22.2 km

Cell value: number of
instances falling in the cell

Methodology

Spatial Signature -- Spatial point patterns -- Examples

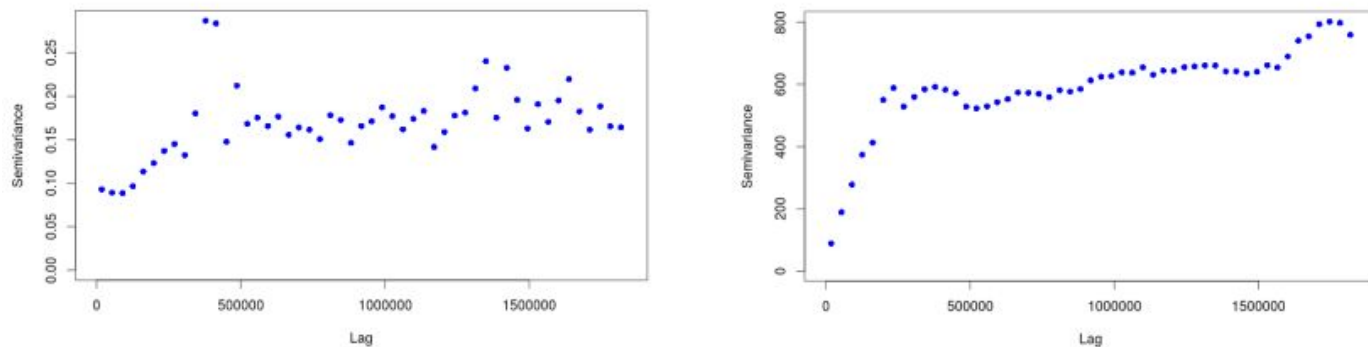


Figure 2: Experimental semivariogram of *Park* (left) and *Dam* (right) from TGN.

Statistics: mean and std. of the semivariance at first, median and last lag distance

Methodology

Spatial Signature -- Spatial Interaction with Other Geographic features

- Population
- Climate
- Road network

Population (LandScan2014)

Population for each feature point



Road Segment (Digital Chart of the World)

Distance to nearest segment for each feature point



- Minimum
- Maximum
- Mean
- Standard deviation

Methodology

Spatial Signature -- Spatial Interaction with Other Geographic features --
Examples

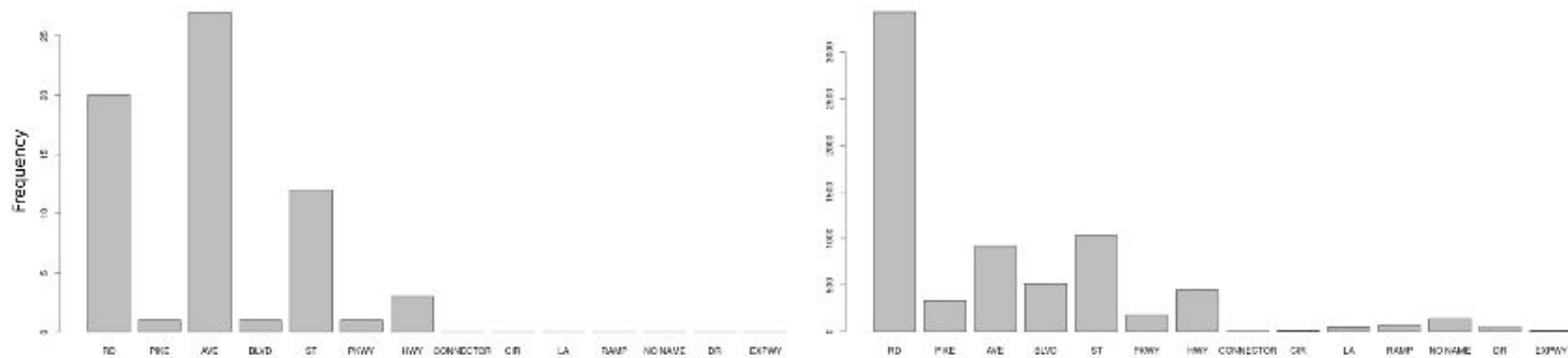


Figure 3: Histogram of road types for *Amusement Park* (left) and *Restaurant* from Google.



Methodology

Spatial Signature -- Place-based statistics

In contrast to spatial statistics, they focus more on describing the ***topological*** and ***hierarchical relations*** between places.

- The number (and entropy) of distinct states (or counties) a place type occurs in;
- The number (and entropy) of adjacent states (or counties) that also contain features of the target type;



Methodology

Spatial Signature -- Place-based statistics -- Examples

- Distinguishing feature types such as Glacier (which occur in eight US-states according to DBpedia) and River (which occur in all states).



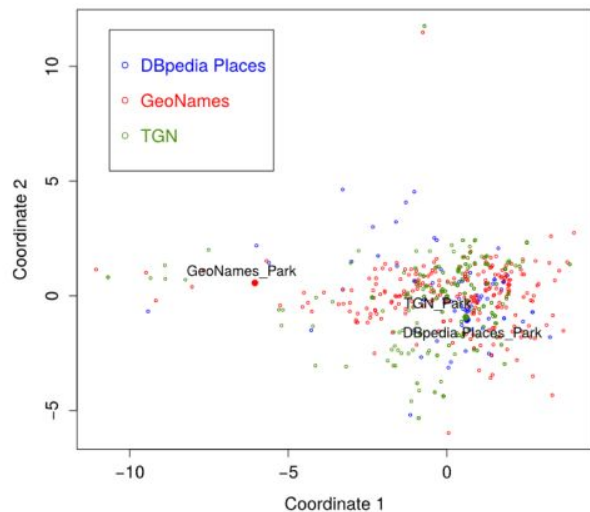
Methodology

Spatial Point Pattern		Spatial Autocorrelations	Spatial Interaction with Other Geographic Features		Place-based statistics	
Local	Intensity	Global Moran's I	Population	min	Number of distinct states (or counties)	
	Mean distance to nearest neighbor			max		
	std. of distance to nearest neighbor			mean	Entropy of states (or counties)	
	Kernel density (range)			std.	Number of adjacent states (or counties) that have the same feature type	
	Kernel density (bandwidth)		Road Network	min of shortest distance		
	Ripley's K (range)			max of shortest distance		
	Ripley's K (mean deviation)			mean of shortest distance	Number of distinct feature types for nearest neighbor	
	std. ellipse (rotation)	std. of shortest distance				
	std. ellipse (std. along x-axis)	Semivariogram (first distance lag)	entropy of nearest road types	Entropy of feature types for nearest neighbor		
	std. ellipse (std. along y-axis)		mean precipitation			
Global	Intensity		Semivariogram (median distance lag)	Climate	std. precipitation	LDA-based approach
		mean temperature max				
	std. temperature max					
	mean temperature min					
	Kernel density (range)	Semivariogram (last distance lag)	std. temperature min		Entropy of the topic distribution	
			mean water vapor pressure			
	std. water vapor pressure					
	Kernel density (bandwidth)					

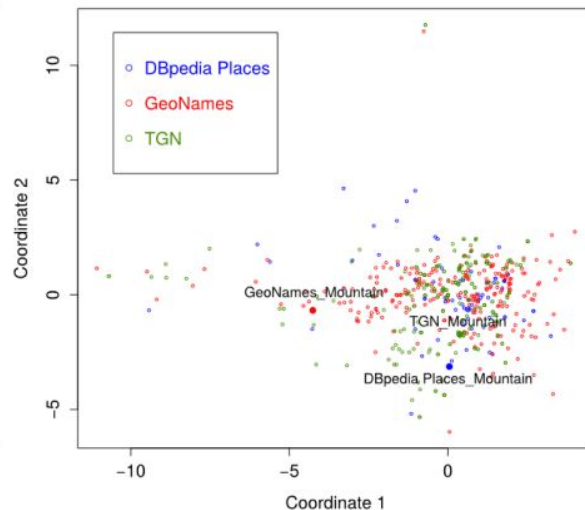
Experiment

1. Similarity of place types

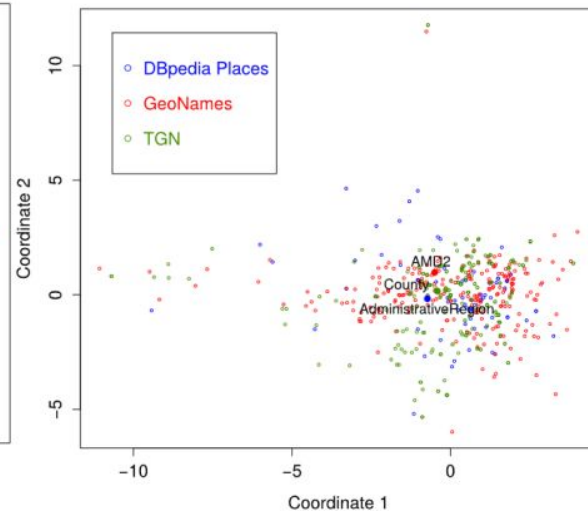
Case 1



Case 2



Case 3



Experiment

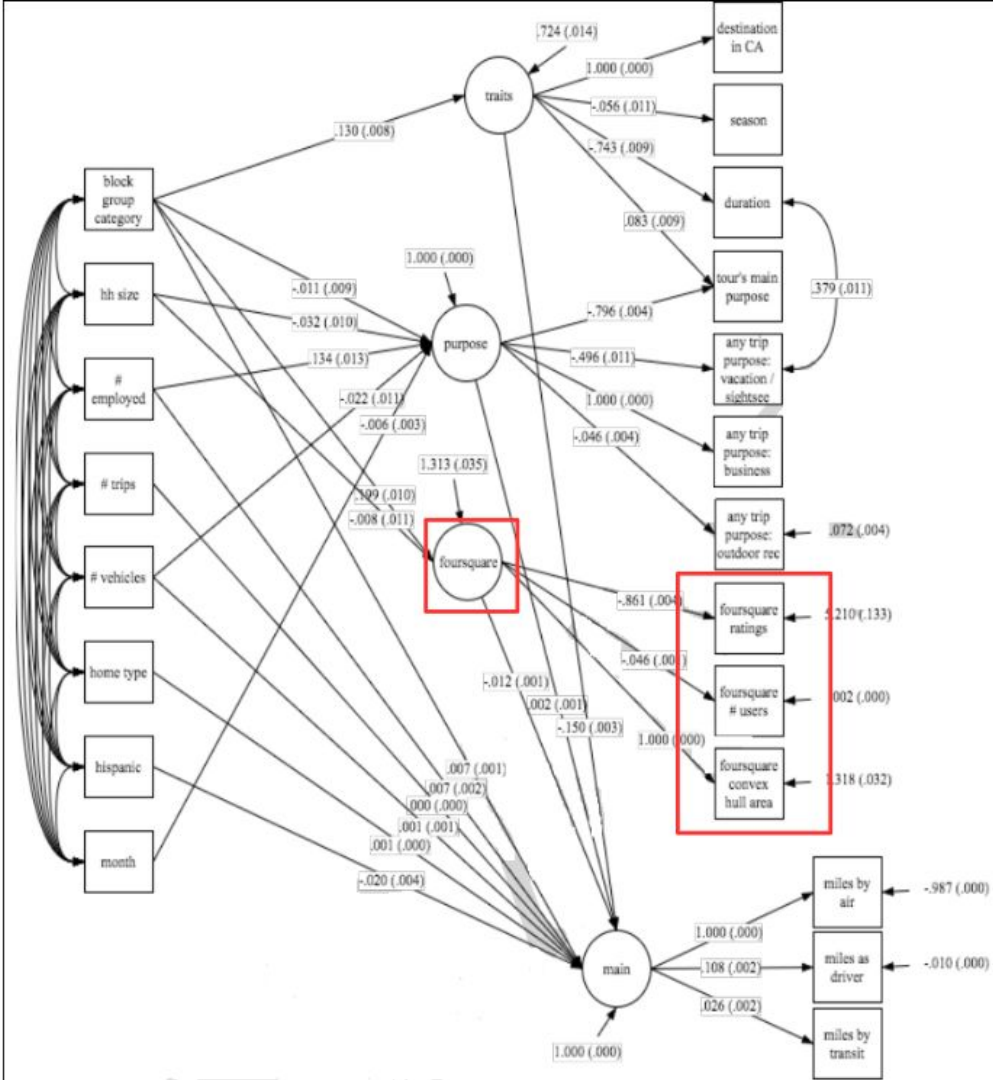
2. Modeling of the LD pattern

- Structural Equation Model

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

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

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





Summary

- External information could be incorporated to understand human travel behaviors;
- Volunteered geographic information has potentials to provide such external information;
- Quantifying the semantic uncertainty of VGI help us to understand places/neighborhoods.



Thanks a lot!

Any questions / comments?