

9-4-2019

Webinar: Social Transportation Analytic Toolbox (STAT) for Transit Networks

Xiaoyue Cathy Liu
University of Utah

Follow this and additional works at: https://pdxscholar.library.pdx.edu/trec_webinar



Part of the [Transportation Commons](#), [Urban Studies Commons](#), and the [Urban Studies and Planning Commons](#)

Let us know how access to this document benefits you.

Recommended Citation

Liu, Xiaoyue Cathy, "Webinar: Social Transportation Analytic Toolbox (STAT) for Transit Networks" (2019). *TREC Webinar Series*. 40.
https://pdxscholar.library.pdx.edu/trec_webinar/40

This Book is brought to you for free and open access. It has been accepted for inclusion in TREC Webinar Series by an authorized administrator of PDXScholar. For more information, please contact pdxscholar@pdx.edu.

Social-Transportation Analytic Toolbox (STAT) for Transit Networks

Xiaoyue Cathy Liu

Department of Civil & Environmental Engineering

University of Utah

Email: cathy.liu@utah.edu

NITC-RR-1080



Xiaoyue Cathy Liu

Shenruoyang Na

Qian Zuo

Ran Wei

Aaron Golub

Liming Wang

Jake Davis



Portland
State
UNIVERSITY

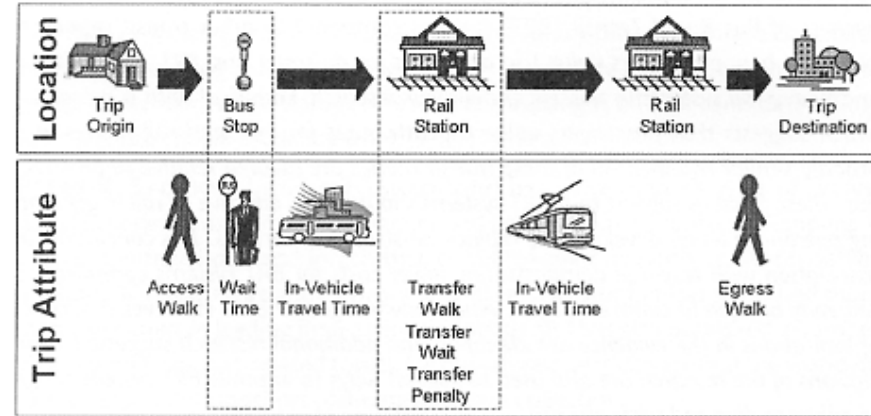
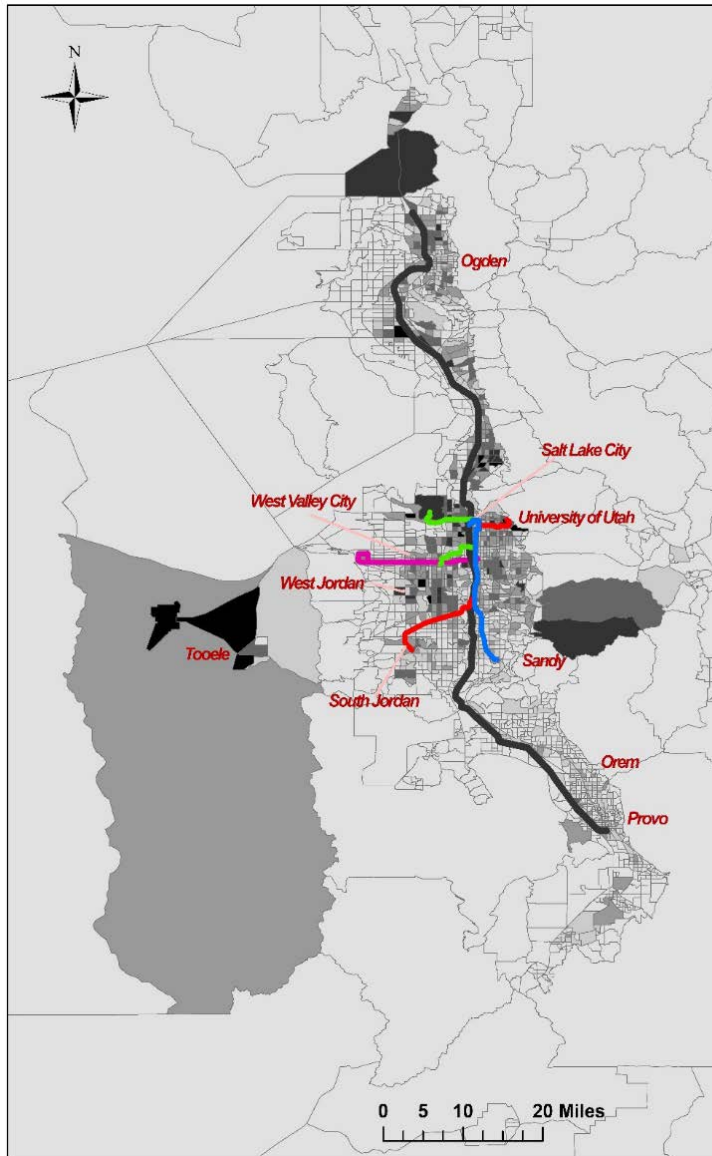
Outline

- ▶ Background
- ▶ Objectives
- ▶ Data
- ▶ Methodology
- ▶ Results
 - STAT Platform
 - Analysis
- ▶ Conclusion

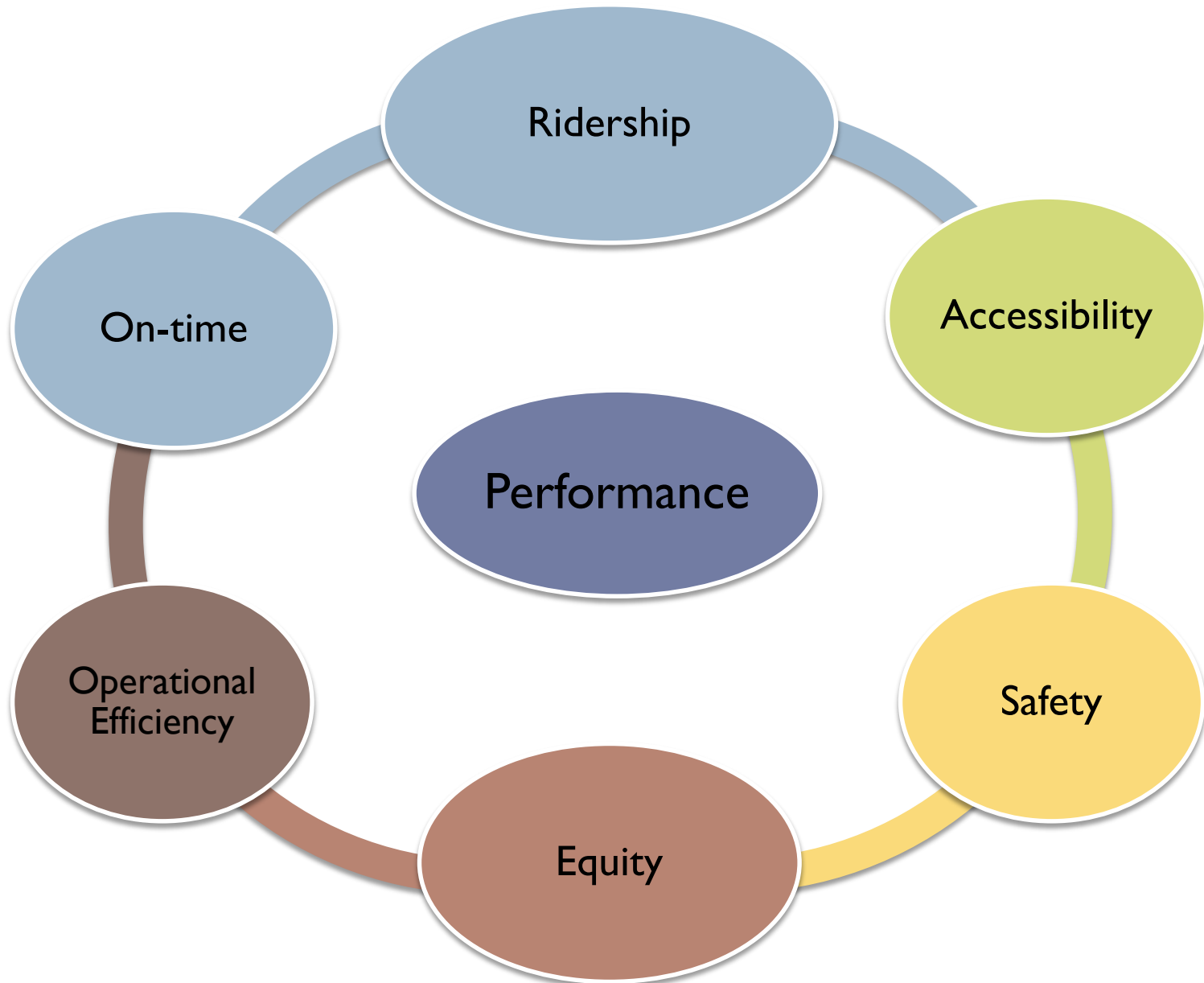
Public Transit System

Wasatch Front Region

UTA's Network



Public Transit Performance Assessment



How to make transit appealing?

- ▶ Regular assessment and improvement of public transit service performance are essential for transit agencies.
 - Identify connectivity gaps
 - Plan for potential investments
- ▶ Evaluating transit service performance is complex and challenging
 - Supply and demand conflict
 - Public perceptions

How to make transit appealing?

- ▶ How agencies feel about the existing transit service vs. transit users' experience?

Can we infer from data?



unTriMet alerts
@unTriMetAlerts

Follow

#trimet MAX Blue Line to Gresham experiencing up to 15 minute delays thru 10:00am from earlier medical issue near Ruby Junction/ E 197t...

9:36 AM - 28 Sep 2018

Name	Date
british_rail_network	Ymtd
agency.txt	17 Au
calendar.txt	17 Au
routes.txt	17 Au
stop_times.txt	17 Au
stops.txt	17 Au
transfers.txt	17 Au
trips.txt	17 Au
british_rail.zip	Ymtd
delijn	25 Oc
agency.txt	21 Oc
calendar_dates.txt	21 Oc
routes.txt	21 Oc
stop_times.txt	21 Oc
stops.txt	21 Oc
trips.txt	21 Oc
delijn.zip	24 Oc
rmsas	24 Oc
agency.txt	3 Au
calendar_dates.txt	7 Au
feed_info.txt	7 Au
routes.txt	7 Au
stop_times.txt	4 Sep
stops.txt	7 Au
trips.txt	7 Au
rmsas.zip	24 Oc
agency.txt	20 Oc
calendar_dates.txt	20 Oc
feed_info.txt	20 Oc
routes.txt	20 Oc
shapes.txt	20 Oc
stop_times.txt	20 Oc
stops.txt	20 Oc

stop_id	stop_code	stop_name	stop_lat	stop_lon	stop_url	location_type	parent_station	wheelchair_boarding
AAP	AAP	Alexandra Palace	51.50973	-0.12873	http://www.nationalrail.co.uk/stations/AAP/details.html	1		1
AAT	AAT	Achanalt	57.48958	-4.91386	http://www.nationalrail.co.uk/stations/AAT/details.html	2		2
ABA	ABA	Aberdare	51.71506	-3.44318	http://www.nationalrail.co.uk/stations/ABA/details.html	1		1
ABC	ABC	Alinabreac	58.18813	-3.78629	http://www.nationalrail.co.uk/stations/ABC/details.html	1		2
ABD	ABD	Aberdeen	57.14169	-2.48969	http://www.nationalrail.co.uk/stations/ABD/details.html	1		1
ABE	ABE	Aber	51.57496	-3.22984	http://www.nationalrail.co.uk/stations/ABE/details.html	1		2
ABP	ABP	Ashurst (Kent) - Rate Face	51.38078	0.15847		1		0
ABM	ABM	Abererch	52.89868	-4.37428	http://www.nationalrail.co.uk/stations/ABM/details.html	1		2
ABW	ABW	Abbey Wood (London)	51.49187	0.12141	http://www.nationalrail.co.uk/stations/ABW/details.html	1		2
ADY	ADY	Ashbury	52.47166	-2.19444	http://www.nationalrail.co.uk/stations/ADY/details.html	1		2
ACB	ACB	Acton Bridge	51.24652	-2.48932	http://www.nationalrail.co.uk/stations/ACB/details.html	1		2
ACC	ACC	Acton Central	51.58872	-0.26297	http://www.nationalrail.co.uk/stations/ACC/details.html	1		2
ACG	ACG	Acorns Green	52.44920	-1.81898	http://www.nationalrail.co.uk/stations/ACG/details.html	1		2
ACH	ACH	Achnashele	57.48289	-5.12384	http://www.nationalrail.co.uk/stations/ACH/details.html	1		2
ACK	ACK	Acklington	55.18718	-1.65184	http://www.nationalrail.co.uk/stations/ACK/details.html	1		1
ACL	ACL	Acle	52.63471	-1.54393	http://www.nationalrail.co.uk/stations/ACL/details.html	1		2
ACN	ACN	Achnasheen	57.57927	-5.87233	http://www.nationalrail.co.uk/stations/ACN/details.html	1		2
ACR	ACR	Accrington	53.75298	-2.36955	http://www.nationalrail.co.uk/stations/ACR/details.html	1		1
ACT	ACT	Ascot	51.48025	-0.67583	http://www.nationalrail.co.uk/stations/ACT/details.html	1		2
ACY	ACY	Abercynon	51.64471	-3.32788	http://www.nationalrail.co.uk/stations/ACY/details.html	1		1
ADC	ADC	Adlington (Cheshire)	53.18597	-2.13356	http://www.nationalrail.co.uk/stations/ADC/details.html	1		1
ADU	ADU	Alderley Park	52.48380	-1.85504	http://www.nationalrail.co.uk/stations/ADU/details.html	1		2
ADK	ADK	Ardwick	53.47157	-2.21432	http://www.nationalrail.co.uk/stations/ADK/details.html	1		2
ADL	ADL	Adlington (Lancs)	53.81326	-2.68387	http://www.nationalrail.co.uk/stations/ADL/details.html	1		1
ADM	ADM	Adisham	51.24128	-1.10018	http://www.nationalrail.co.uk/stations/ADM/details.html	1		1
ADN	ADN	Ardrossan Town	55.63978	-4.81267	http://www.nationalrail.co.uk/stations/ADN/details.html	1		1
ADR	ADR	Airdrie	55.86377	-3.98258	http://www.nationalrail.co.uk/stations/ADR/details.html	1		1
ADP	ADP	Ardrossan Harbour	55.63085	-4.82111	http://www.nationalrail.co.uk/stations/ADP/details.html	1		1
ADV	ADV	Andover	51.21355	-1.49224	http://www.nationalrail.co.uk/stations/ADV/details.html	1		1
ADW	ADW	Adlestree	51.55826	-1.82852	http://www.nationalrail.co.uk/stations/ADW/details.html	1		2
AFK	AFK	Ashford International	51.14371	0.87621	http://www.nationalrail.co.uk/stations/AFK/details.html	1		1
AFS	AFS	Ashford (Surrey)	51.43631	-0.46886	http://www.nationalrail.co.uk/stations/AFS/details.html	1		2
AFV	AFV	Arsdell & Fairbairn	53.74547	-2.59384	http://www.nationalrail.co.uk/stations/AFV/details.html	1		1
AGL	AGL	Abergele & Pensarn	53.29459	-3.58263	http://www.nationalrail.co.uk/stations/AGL/details.html	1		1
AGR	AGR	Angel Road	51.81241	-0.84878	http://www.nationalrail.co.uk/stations/AGR/details.html	1		2
AGS	AGS	Argyle Street	55.89729	-2.24869	http://www.nationalrail.co.uk/stations/AGS/details.html	1		2
AGT	AGT	Aldrington	58.83638	-0.18382	http://www.nationalrail.co.uk/stations/AGT/details.html	1		1
AGV	AGV	Aberavenny	51.81689	-3.48986	http://www.nationalrail.co.uk/stations/AGV/details.html	1		2
AGW	AGW	Achtod	53.13187	-0.18877	http://www.nationalrail.co.uk/stations/AGW/details.html	1		1
ANM	ANM	Ashton-under-Lyne	53.49128	-2.89431	http://www.nationalrail.co.uk/stations/ANM/details.html	1		1
ANB	ANB	Ashurst (Kent)	51.12866	0.15266	http://www.nationalrail.co.uk/stations/ANB/details.html	1		2
ANT	ANT	Alderbury	51.24614	-0.75288	http://www.nationalrail.co.uk/stations/ANT/details.html	1		2
ANV	ANV	Ash Vale	51.27225	-0.72165	http://www.nationalrail.co.uk/stations/ANV/details.html	1		1
ANB	ANB	Ashted	51.24614	-0.75288	http://www.nationalrail.co.uk/stations/ANB/details.html	1		2
AIN	AIN	Aintree	53.47392	-2.95629	http://www.nationalrail.co.uk/stations/AIN/details.html	1		1
AIR	AIR	Airbles	55.78282	-3.99418	http://www.nationalrail.co.uk/stations/AIR/details.html	1		1
ALB	ALB	Albrighton	52.63796	-2.16891	http://www.nationalrail.co.uk/stations/ALB/details.html	1		2
ALD	ALD	Alderley Edge	53.38380	-2.23680	http://www.nationalrail.co.uk/stations/ALD/details.html	1		1
ALF	ALF	Alfreton	53.18844	-1.36969	http://www.nationalrail.co.uk/stations/ALF/details.html	1		2

train
busstop
station
line
riding
trax
wrong
bikes
runner
delays
trains
front
crossing
central
northbound
murray
wait
signs
salt
north
late
runner
front

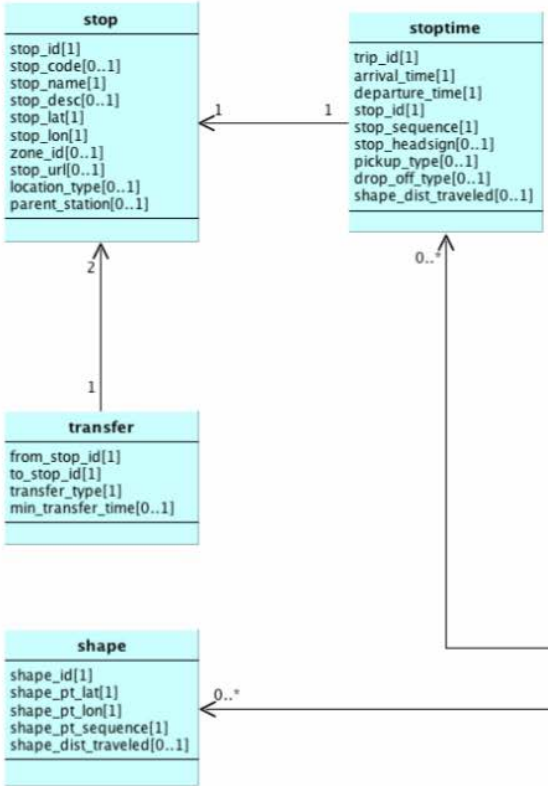
Research Objectives

- ▶ Leverage open transportation data to discover transit service connectivity and users' opinion
 - ▶ GTFS
 - ▶ Twitter
- ▶ Build an open-source *social-transportation analytic toolbox (STAT)* to explore fundamental patterns of interactions between users and transit services

GTFS – General Transit Feed Specification

```

trip_id,arrival_time,departure_time,stop_id,stop_sequence,stop_headsign,pickup_time,drop_off_type,shape_dist_traveled
1966787,6:25:00,6:25:00,21168,1,,0,0,
1966787,6:28:45,6:28:45,5479,2,,0,0,
1966787,6:29:13,6:29:13,13635,3,,0,0,
1966787,6:29:54,6:29:54,5481,4,,0,0,
1966787,6:30:45,6:30:45,13105,5,,0,0,
1966787,6:32:00,6:32:00,5504,6,,0,0,
1966787,6:32:41,6:32:41,13192,7,,0,0,
1966787,6:33:38,6:33:38,15182,8,,0,0,
1966787,6:34:12,6:34:12,15183,9,,0,0,
1966787,6:34:38,6:34:38,15184,10,,0,0,
1966787,6:35:04,6:35:04,15185,11,,0,0,
1966787,6:35:24,6:35:24,15186,12,,0,0,
1966787,6:35:44,6:35:44,15187,13,,0,0,
1966787,6:36:24,6:36:24,15188,14,,0,0,
1966787,6:37:05,6:37:05,15189,15,,0,0,
1966787,6:37:34,6:37:34,15190,16,,0,0,
1966787,6:38:00,6:38:00,3569,17,,0,0,
1966787,6:38:18,6:38:18,15191,18,,0,0,
1966787,6:38:35,6:38:35,15205,19,,0,0,
1966787,6:38:49,6:38:49,15206,20,,0,0,
1966787,6:39:32,6:39:32,21051,21,,0,0,
1966787,6:40:00,6:40:00,2773,22,,0,0,
1966787,6:40:14,6:40:14,2774,23,,0,0,
1966787,6:40:27,6:40:27,15209,24,,0,0,
1966787,6:40:49,6:40:49,15210,25,,0,0,
1966787,6:41:10,6:41:10,15211,26,,0,0,
1966787,6:41:29,6:41:29,2778,27,,0,0,
1966787,6:42:05,6:42:05,15212,28,,0,0,
1966787,6:42:31,6:42:31,21284,29,,0,0,
1966787,6:44:00,6:44:00,2780,30,,0,0,
1966787,6:45:01,6:45:01,15213,31,,0,0,
1966787,6:45:21,6:45:21,15214,32,,0,0,
1966787,6:45:42,6:45:42,3583,33,,0,0,
1966787,6:52:14,6:52:14,15195,34,,0,0,
1966787,6:53:00,6:53:00,1022,35,,0,0,
1966788,6:55:00,6:55:00,21168,1,,0,0,
1966788,6:59:50,6:59:50,5479,2,,0,0,
1966788,7:00:26,7:00:26,13635,3,,0,0,
1966788,7:01:19,7:01:19,5481,4,,0,0,
1966788,7:02:25,7:02:25,13105,5,,0,0,
1966788,7:04:00,7:04:00,5504,6,,0,0,
1966788,7:04:48,7:04:48,13192,7,,0,0,
1966788,7:05:54,7:05:54,15182,8,,0,0,
1966788,7:06:34,7:06:34,15183,9,,0,0,
1966788,7:07:04,7:07:04,15184,10,,0,0,
1966788,7:07:35,7:07:35,15185,11,,0,0,
1966788,7:07:58,7:07:58,15186,12,,0,0,
1966788,7:08:22,7:08:22,15187,13,,0,0,
1966788,7:09:08,7:09:08,15188,14,,0,0,
1966788,7:09:56,7:09:56,15189,15,,0,0,
1966788,7:10:30,7:10:30,15190,16,,0,0,
1966788,7:11:00,7:11:00,3569,17,,0,0,
1966788,7:11:24,7:11:24,15191,18,,0,0,
    
```



Twitter

- ▶ Retrieve from twitter.com
- ▶ Advantage:
 - Big data
- ▶ Disadvantage:
 - Only 1% of total volume
 - At rough geographic level
 - Cannot provide contextual inferences - most studies only use density as a surrogate measure for population

The image shows a screenshot of a Twitter profile for 'Civil Engineering @UofUCvEEN'. The profile header includes the name, handle, and statistics: 485 Tweets, 92 Following, 7,991 Followers, 44 Likes, and 1 List. The main content area displays a series of tweets and retweets. The first tweet is a link to a research paper on microbial resource management. The second tweet is a promotional post for a publication titled 'Biking islands in cities: An analysis combining bike trajectory and pe...'. The third tweet is a retweet from NITC @NITC_UTC, dated August 20, which promotes a webinar titled 'Social Transportation Analytic Toolbox (STAT) for Transit Networks' held on September 4, 2019. The webinar poster features a map of the Portland area with various transit routes and stations marked. The fourth tweet is a promotional poster for an 'Engineering CLUB RUSH' event on August 29, 2019, from 11:30 AM to 2:00 PM, held at the Web Lower Patio, with a 'free pizza!!' offer. The profile footer shows a retweet from 'U of U Water Center @uWaterCenter' dated August 19.

Public Transit Accessibility (PTA)

▶ Transit Accessibility

- The ease of travel for an individual to reach a desired destination by public transport
- Helps transit agencies and planners identify areas in need of transit service improvements and prioritize transit investments

Challenges: heterogeneous data, scale

Measure transit accessibility

Why accessibility matters?

Missing temporal dimension -> only focus on specific time-of-day -> overly optimistic evaluation

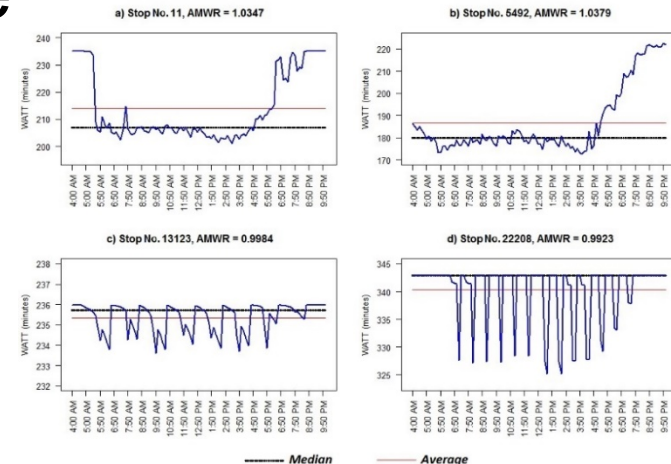
Public Transit Accessibility (PTA)

- ▶ WATT as PTA Measure (weighted average travel time)

- ▶
$$WATT_{i,t} = \frac{\sum_{j=1}^J O_j * tt_{ij,t}}{\sum_{j=1}^J O_j}, \quad j = 1, 2, \dots, J, \quad i \in J$$

- ▶ $WATT_{i,t}$ is the weighted average travel time of stop i at departure time t
- ▶ O_j is the number of opportunities available at stop j
- ▶ $tt_{ij,t}$ is the travel time from stop i to stop j at departure time t
- ▶ J is the total number of stops

- ▶ WATT is of **Travel time** Nature

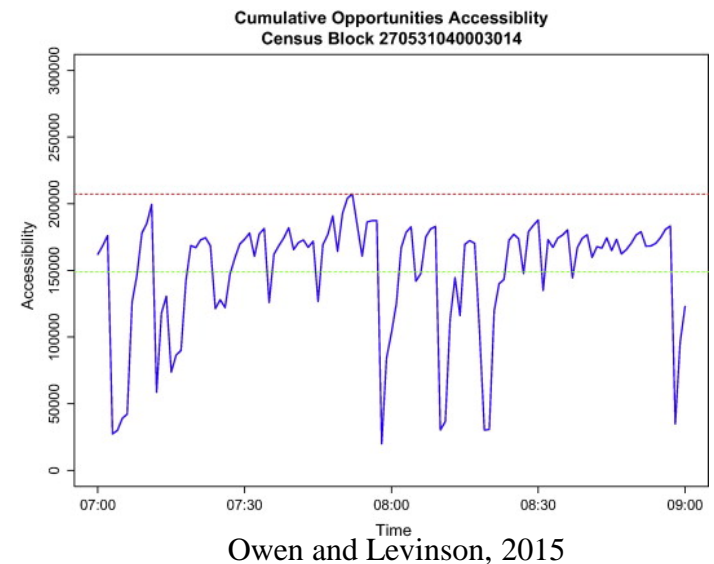


Public Transit Accessibility (PTA)

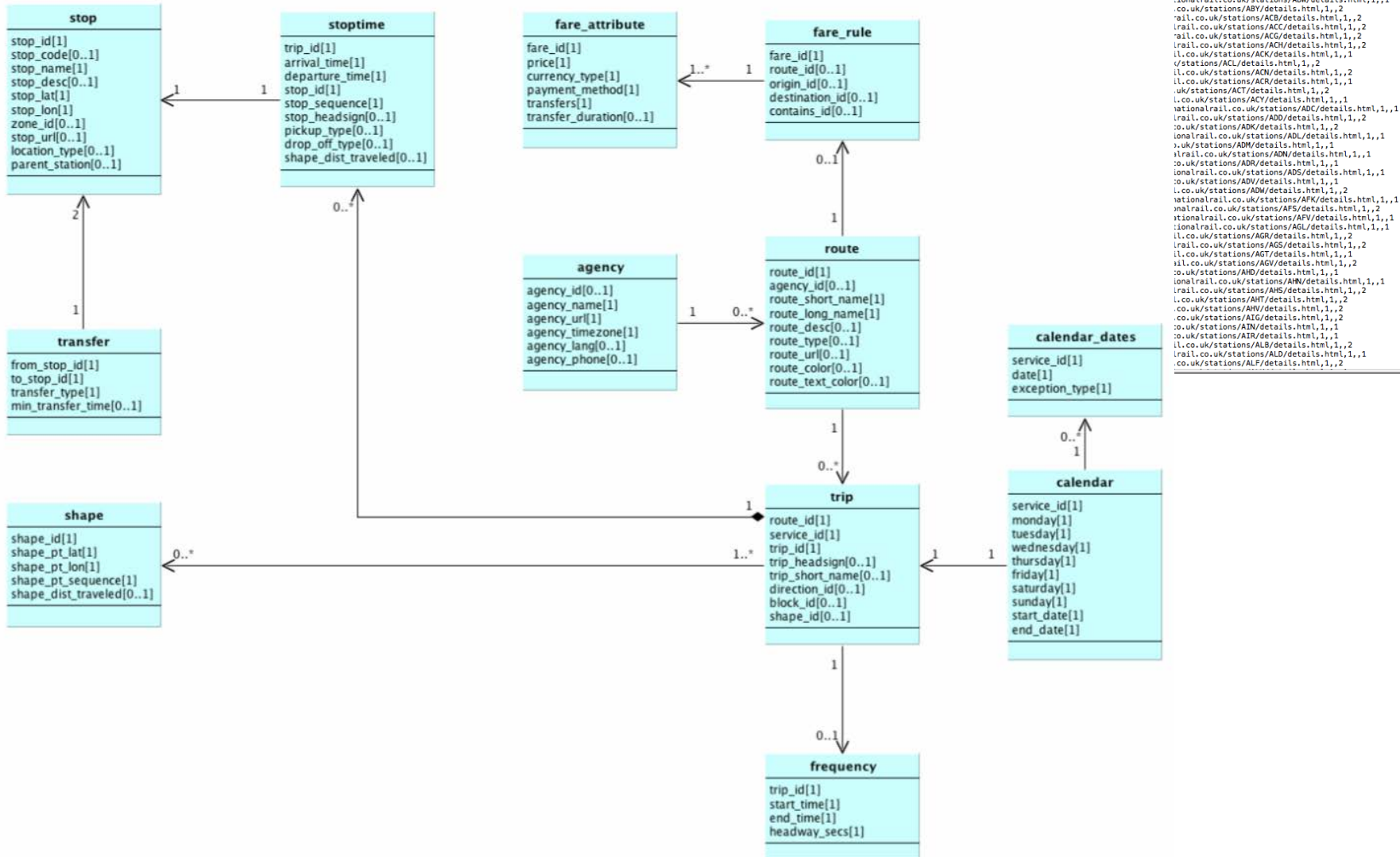
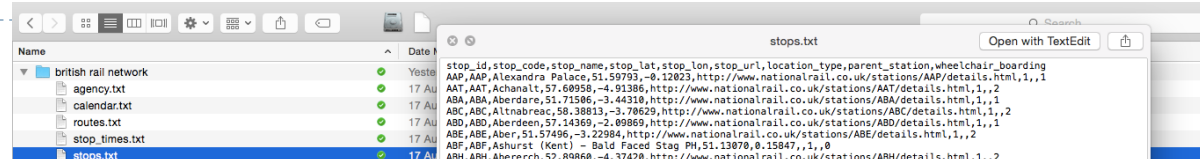
Temporal Fluctuation of PTA

- ▶ PTA for specific time-of-day (e.g. 8:00 AM)
- ▶ Ignoring the temporal fluctuation of PTA
- ▶ Biased results
- ▶ Solution: Travel Times for all ODs must be calculated for all times of day

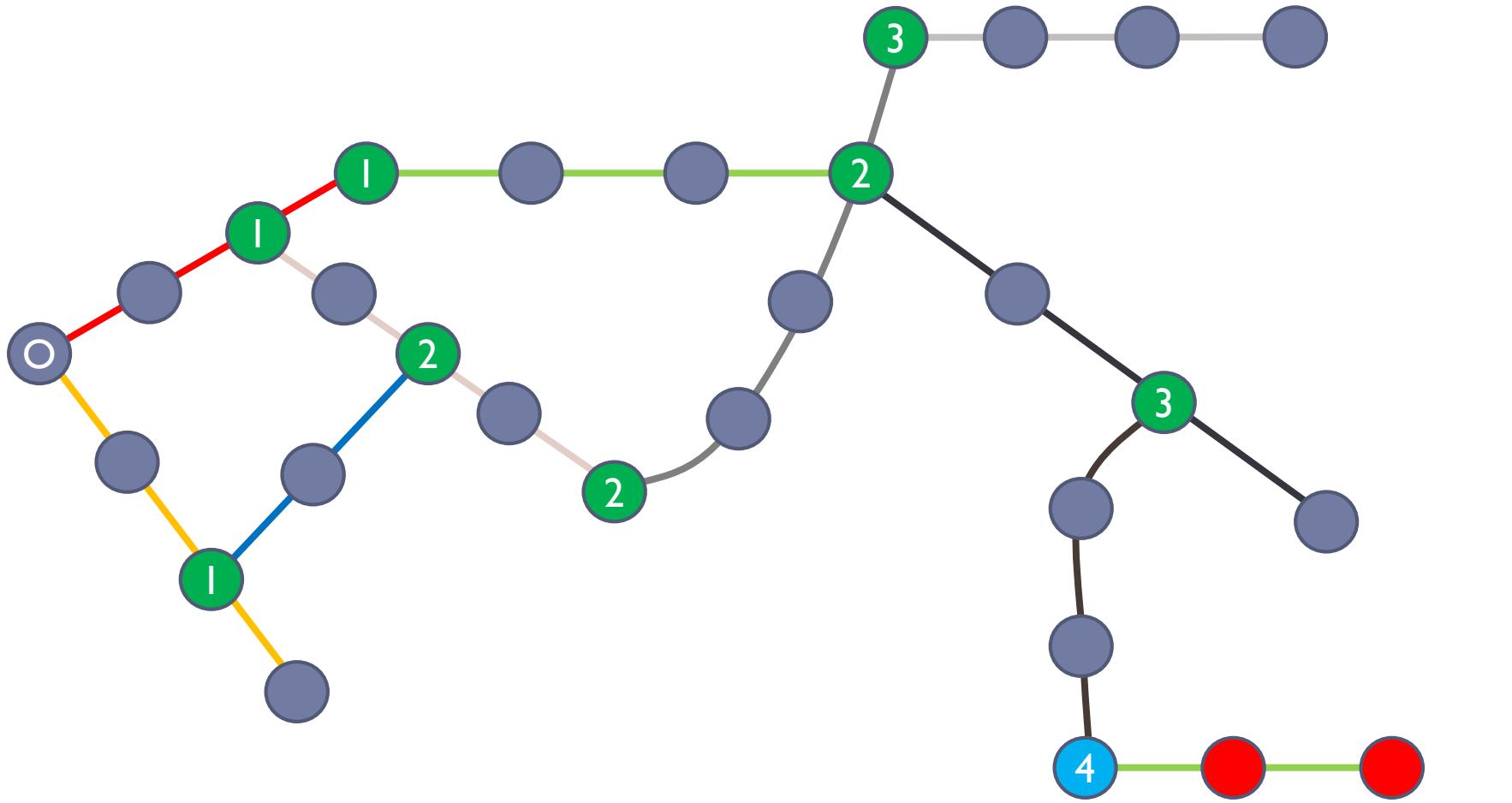
Contributions:
Capture the dynamics of WATT
Identify PTA consistency



Public Transit Accessibility (PTA) - GTFS



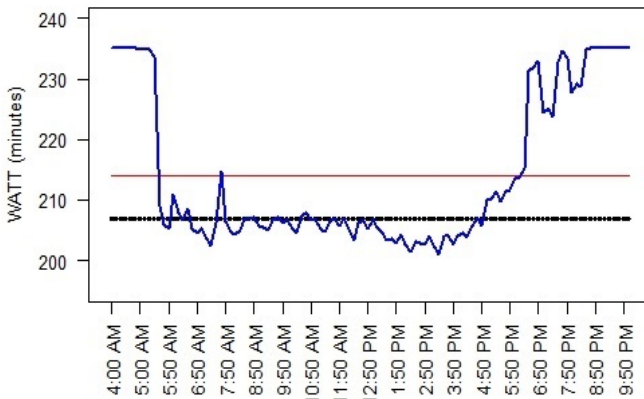
Dynamic PTA Algorithm



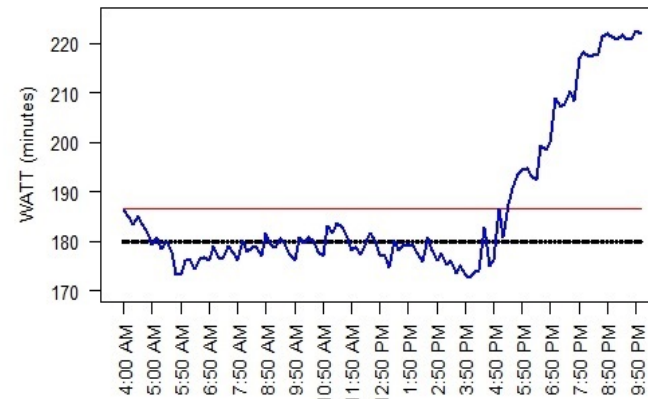
Public Transit Accessibility (PTA)

Quality of service is affected by the probability of WATT for each random departure being closer to minimum as compared to maximum.

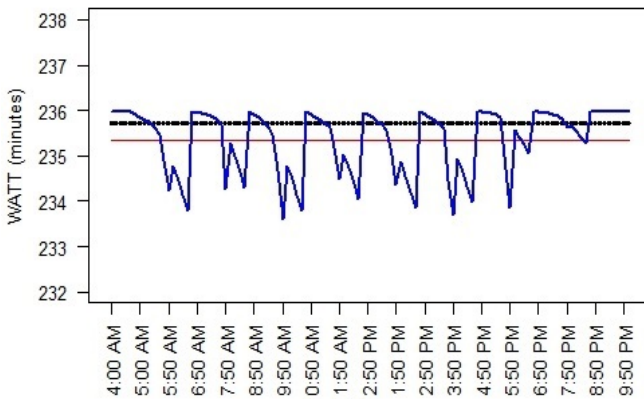
a) Stop No. 11, AMWR = 1.0347



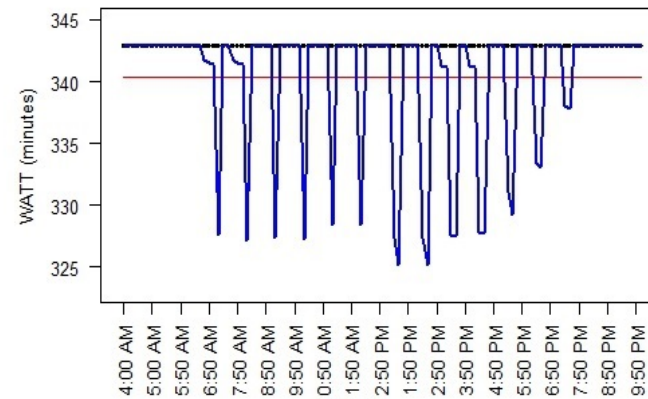
b) Stop No. 5492, AMWR = 1.0379



c) Stop No. 13123, AMWR = 0.9984



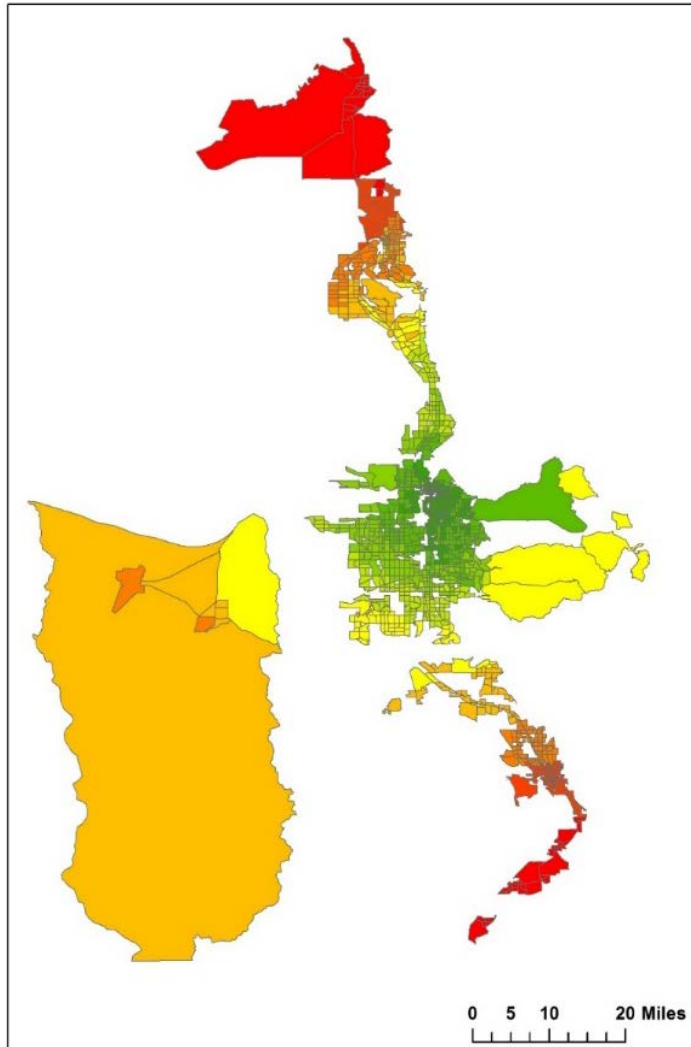
d) Stop No. 22208, AMWR = 0.9923



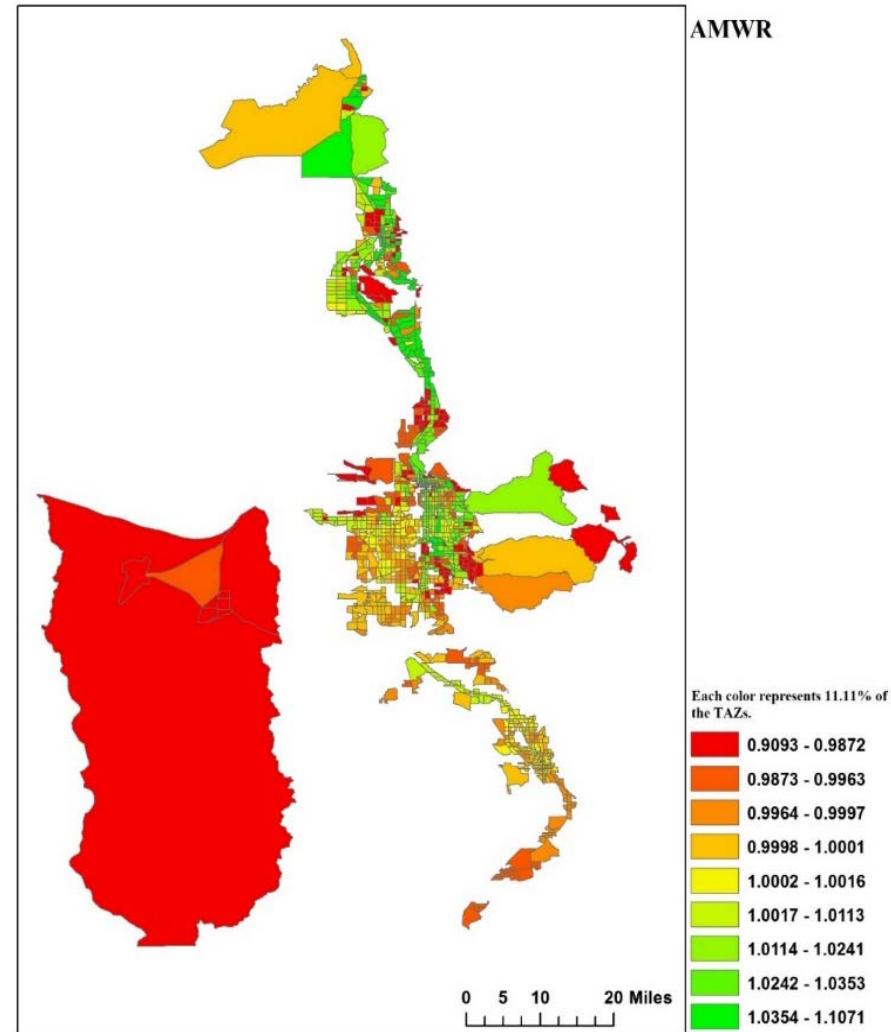
----- Median

----- Average

Public Transit Accessibility (PTA)



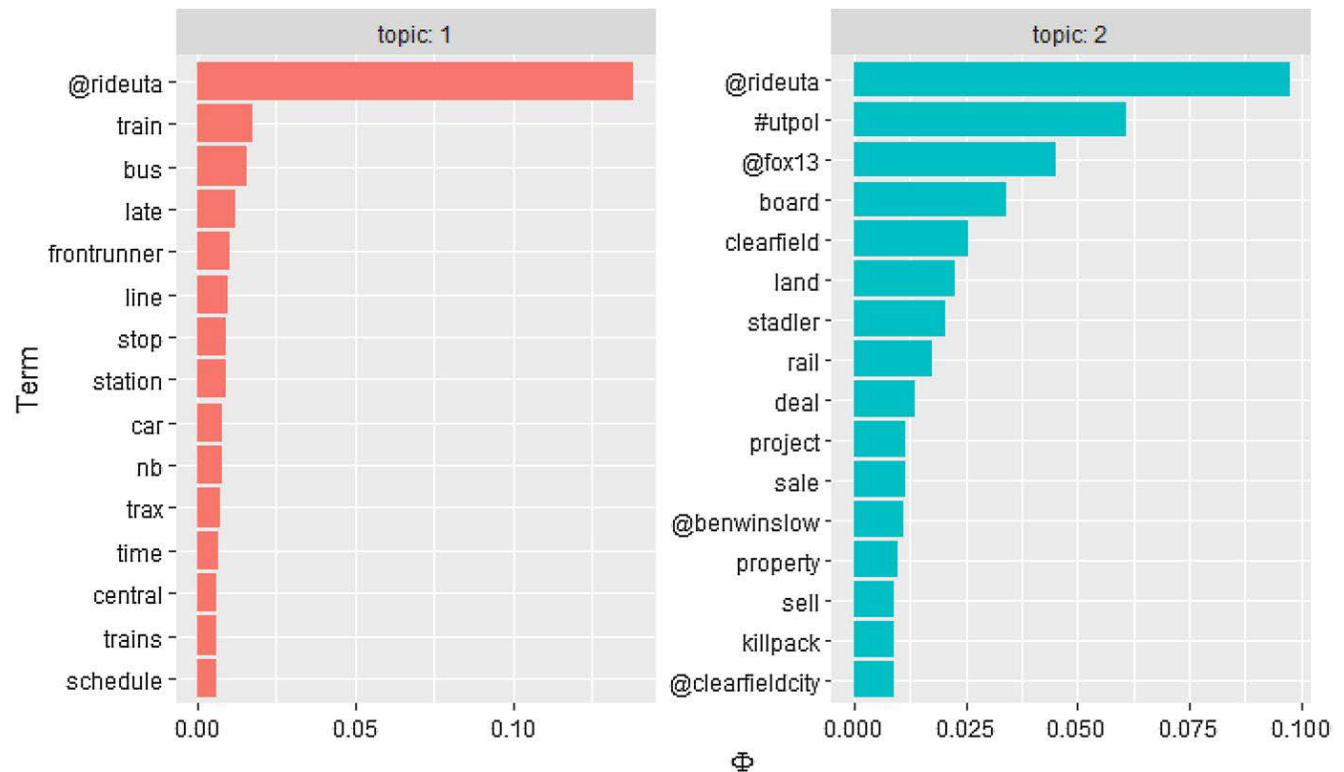
WATT by Public Transit



AMWR

Twitter Data Acquisition and Processing

- ▶ Classify negative semantic about transit
 - User complaints
 - Delay info by transit agencies
- ▶ Information extraction to extract geo information



Classification Model

- ▶ Used python library to retrieve more eligible data
 - ▶ id: a unique number describes the tweet
 - ▶ permalink: the URL link of the tweet
 - ▶ username: the user who publishes the tweet
 - ▶ text: the content of the tweet
 - ▶ date: date the user tweets
 - ▶ retweets: number of retweets
 - ▶ favorites: number of favorites
 - ▶ mentions: mentioned user
 - ▶ hashtags: hashtags of the tweet
 - ▶ geo: geo info of the tweet

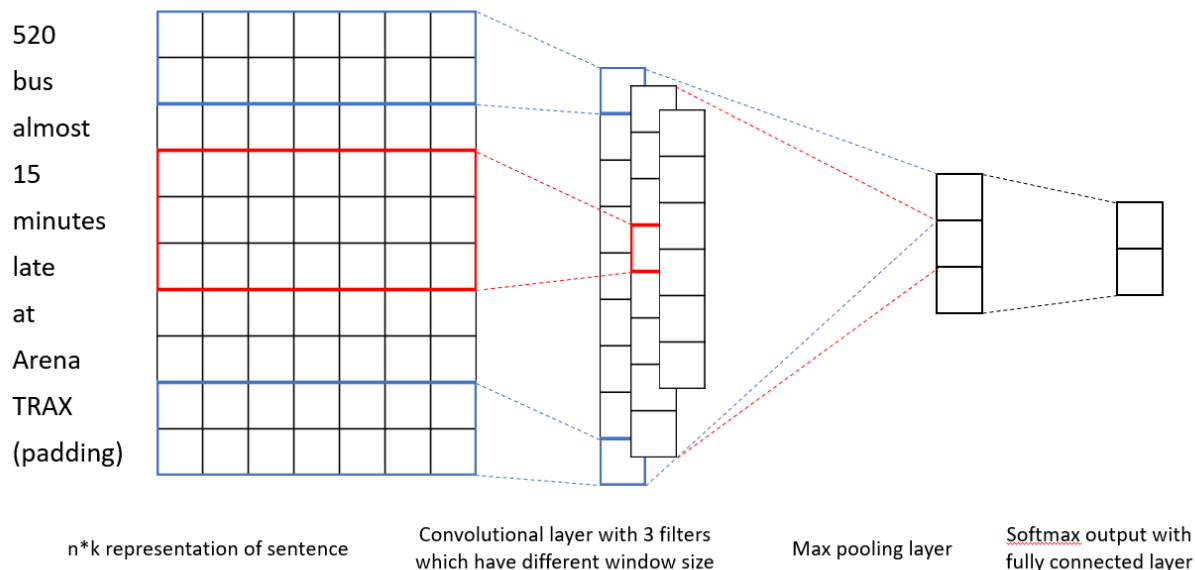
Classification Model (Continued)

▶ Set user-defined criteria

- ▶ setUsername (): An optional specific username(s) from a twitter account (with or without "@").
- ▶ setSince ("yyyy-mm-dd"): A lower bound date (UTC) to restrict search.
- ▶ setUntil ("yyyy-mm-dd"): An upper bound date (not included) to restrict search.
- ▶ setQuerySearch (str): A query text to be matched.
- ▶ setTopTweets (bool): If True only the Top Tweets will be retrieved.
- ▶ setNear(str): A reference location area from where tweets were generated.
- ▶ setWithin (str): A distance radius from "near" location (e.g. 15mi).
- ▶ setMaxTweets (int): The maximum number of tweets to be retrieved. If this number is unsetted or lower than 1 all possible tweets will be retrieved.

Classification Model (Continued)

- ▶ Training data: 15,000 tweets from “@rideuta”
- ▶ Label “complaints” if the tweet contains words “delay” “early” “on time”
- ▶ Use skip-gram model to convert words to vectors
- ▶ CNN to distinguish complaints vs. non-complaints



Information Extraction Method

- ▶ Detect geolocation info of tweets

- Entity detection

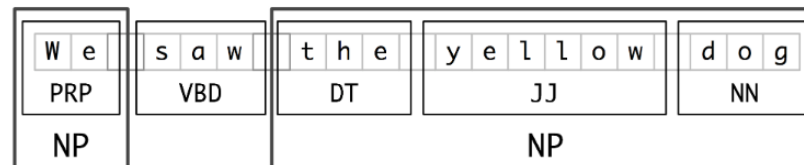
- ▶ Noun phrase (NP) chunking

- Tag all noun phrases in tweets

- Employed the BI-LST-CRF model

- Training data

<https://www.clips.uantwerpen.be/conll2000/chunking/>



- ▶ Measure the similarity between NP and the entities in the gazetteer

Similarity measurement

▶ Edit distance

- ▶ Given two strings str1 and str2 and below operations that can be performed on str1. Find minimum number of edits (operations) required to convert 'str1' into 'str2'.

- Insert
- Remove
- Replace

5	0	0	S
5	0	3	S

▶ Word vector similarity

- ▶ Calculate the Cosine distance between the word vector

If the similarity is larger than some threshold, we consider such Noun phrase is matched with that entity in the gazetteer.

Tweets retrieved

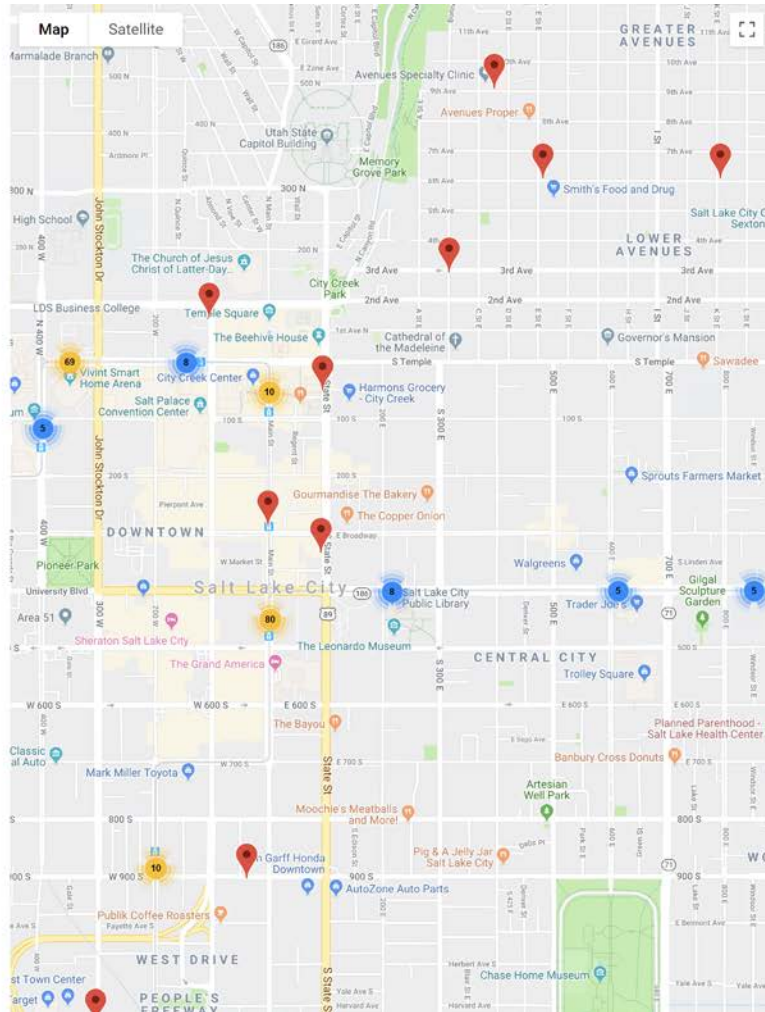
▶ SLC

- ▶ # tweets 15,000~ (training data)
- ▶ Date range 2018-02-21 to 2018-09-29
- ▶ 1800~ – negative semantics

▶ TriMet

- ▶ # tweets 15,000~
- ▶ Date range 2018-06-08 to 2018-09-29
- ▶ 600~ – negative semantics

Tweets geomapping



- Elaine Taylor** @elainetaylor · Apr 3, 2018
Replying to @RideUTA
Library
- UTA** @RideUTA
Thanks. We confirmed through video playback that the NB train scheduled to depart Library at 11:20 am departed on time. We recommend being at a station or stop 5-7 minutes earlier than the scheduled time to ensure you're able to catch the train/bus you want without issue.
7:08 PM · Apr 3, 2018
[See UTA's other Tweets](#)
- UTA** @RideUTA · Mar 29, 2018
Replying to @RideUTA
TRAX Update 8:50 pm: Expect 20-25 min delay between 900 South-Medical on NB Red Line.
Expect 10-15 min delay between Millcreek-900 South on NB Red Line.
There will be one-car SB Red Line trains.
- UTA** @RideUTA
TRAX Update 9:02 pm: Expect 20-25 min delay between 900 East-Medical on NB Red Line.
Expect 10-15 min delay between Library-900 East on NB Red Line.
9:02 PM · Mar 29, 2018
[See UTA's other Tweets](#)
- UTA** @RideUTA
TRAX Alert 2:45 pm: Due to a non-UTA incident blocking the rail, expect delays of about 10 minutes on NB Red Line from Library - Medical Center and on SB Red Line from Stadium - Daybreak Parkway.
2:48 PM · Mar 15, 2018
[See UTA's other Tweets](#)
- Ben Winslow** @BenWinslow · Feb 22, 2018
Replying to @BenWinslow
"I say come to Utah because we are safe," Ruzicka counters to the ads targeting the state.
She says Utah should pay \$\$ to advertise to people to use ride-shares. @fox13 #utpol #utleg

Examples



Clare
@soontobesetter

@RideUTA where is the 7:12 at midvale center to slc.
Seeing no delays but also no train...

7:17 AM · Feb 22, 2018 · [Twitter for iPhone](#)

Matched Stop	Date	Lat,lon
midvale center station	2/22/2018 14:17	40.61054 -111.893
Link : https://twitter.com/soontobesetter/status/966678326176722944		



unTriMet alerts
@unTriMetAlerts

#trimet MAX Blue Line westbound is delayed up to 30
minutes due to an earlier fire near Gateway TC.

4:36 PM · Aug 14, 2018 · [TriMet alerts \(unofficial feed\)](#)

Matched Stop	Date	Lat,lon
gateway transit center	8/14/2018 22:36	45.53081 -122.563
Link : https://twitter.com/unTriMetAlerts/status/1029496898887999488		

STAT Platform

▶ Visualization:

➤ Google Map API

▶ Marker Clustering

<https://developers.google.com/maps/documentation/javascript/marker-clustering>

➤ Twitter API

▶ Embedded Tweets

<https://developer.twitter.com/en/docs/twitter-for-websites/embedded-tweets/overview>

➤ ArcGIS Online API

▶ Web Maps

<https://developers.arcgis.com/documentation/core-concepts/web-maps/>

▶ Data processing:

➤ Machine learning

▶ Tensorflow <https://www.tensorflow.org/>

▶ Anago <https://github.com/Hironsan/anago>

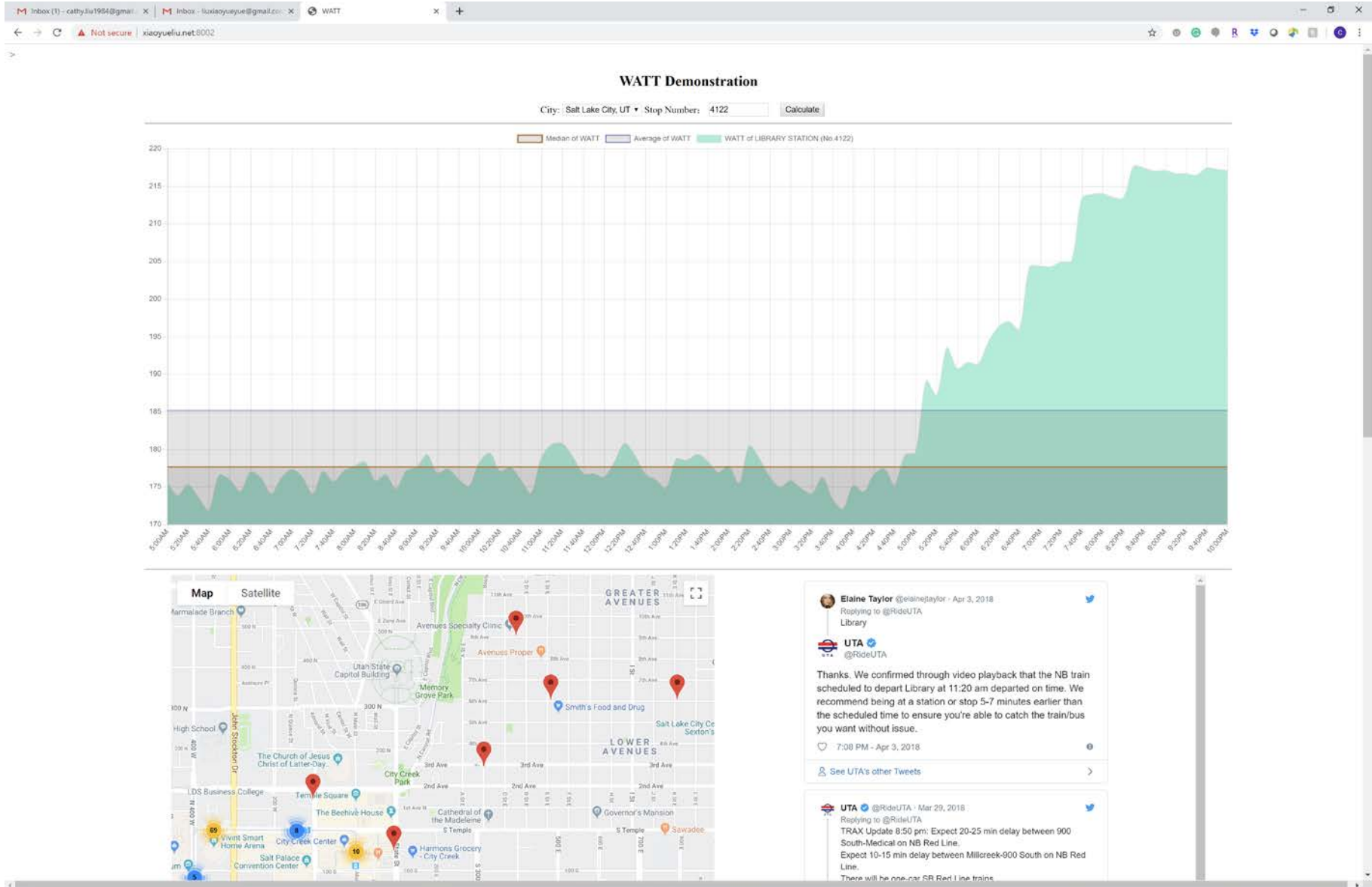
➤ Accessing tweets

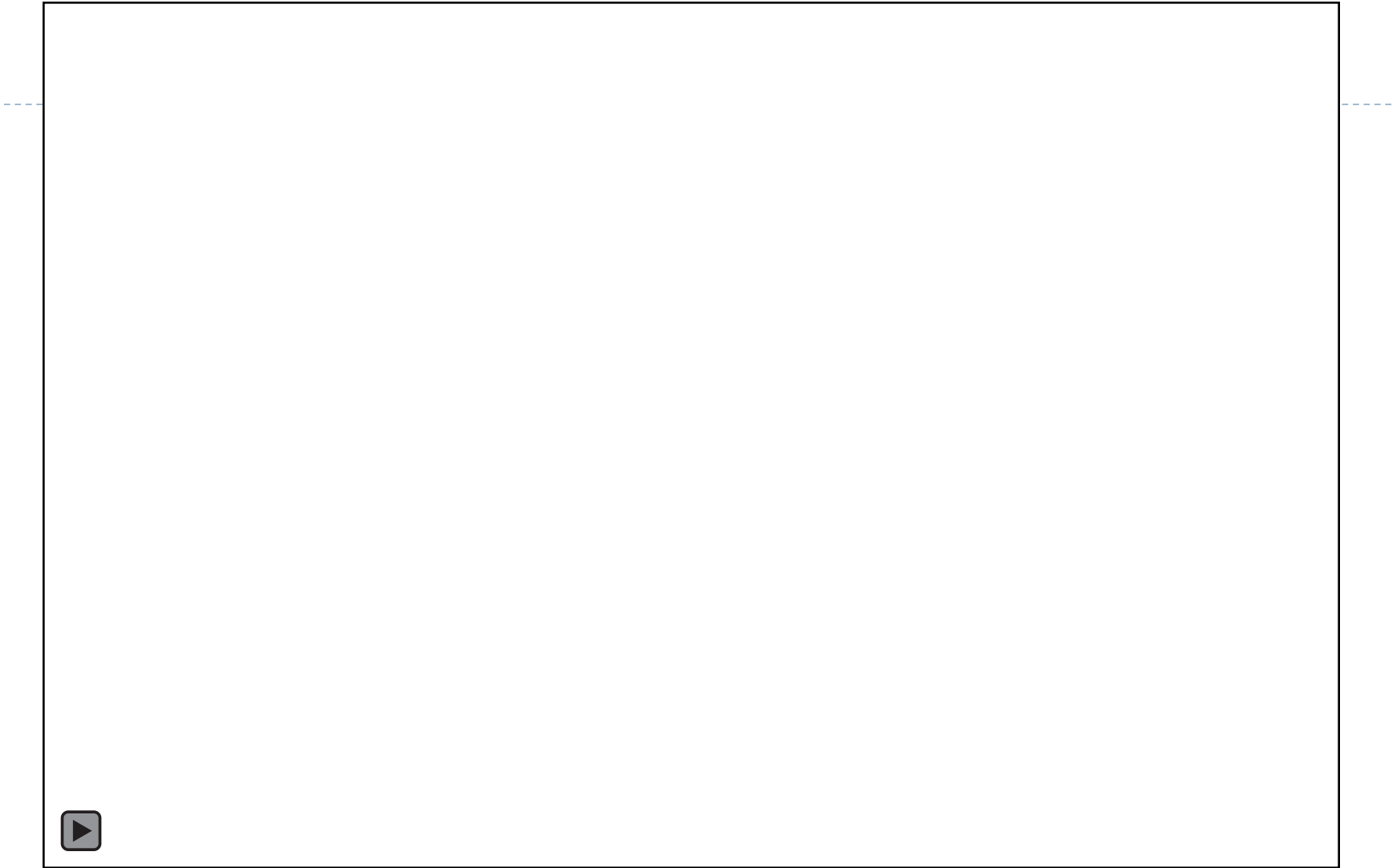
<https://github.com/Mottl/GetOldTweets3>

▶ Web Framework:

➤ Node.js

STAT Platform



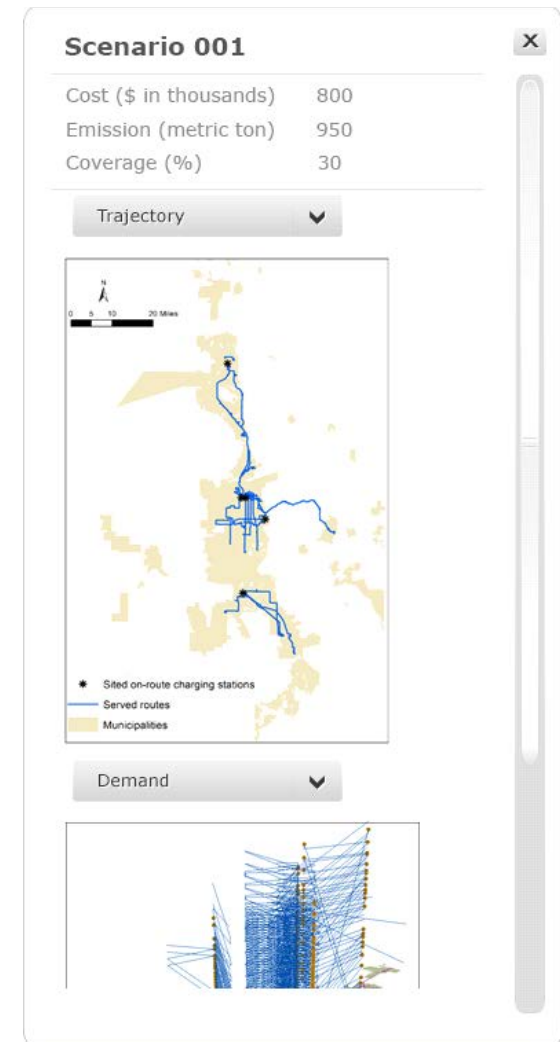
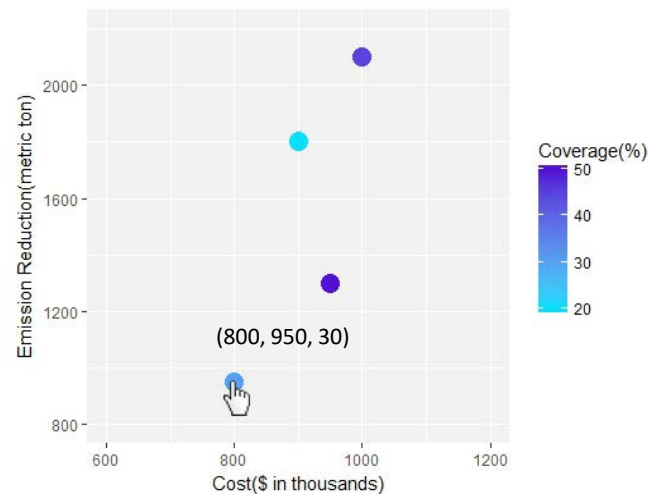


Take-away highlights

- ▶ A proof-of-concept framework to allow users/transit agencies to explore the use and integration of new open transportation data, and discover the interactive pattern between what the transit service is offering vs. what the users are experiencing
 - ▶ Temporal distribution of transit stop's accessibility
 - ▶ Transit stop positioning in Google Maps with geomapped tweets
 - ▶ Overall accessibility visualization in ArcGIS at TAZ level
- ▶ Implemented on UTA and TriMet's networks

Ongoing NITC research

► Electric Bus Deployment Framework



Thanks!

Questions or Comments?

Xiaoyue Cathy Liu
cathy.liu@utah.edu