Cross-Country Analysis of Location-based Sentiment in User-generated Text

Chris Stelzmüller (Johannes Kepler Universität Linz)

Sebastian Tanzer (Johannes Kepler Universität Linz)

Markus Schedl (Johannes Kepler Universität Linz & Linz Institute of Technology)

JⱯU

The Problem We Are Trying to Solve

- Urban planners need data about how people interact within their cities
- Current data collection depends a lot on surveys
- Survey results provide only a snapshot
- We want to provide a continuous stream of how citizens interact with a city's infrastructure
- Current alternatives:
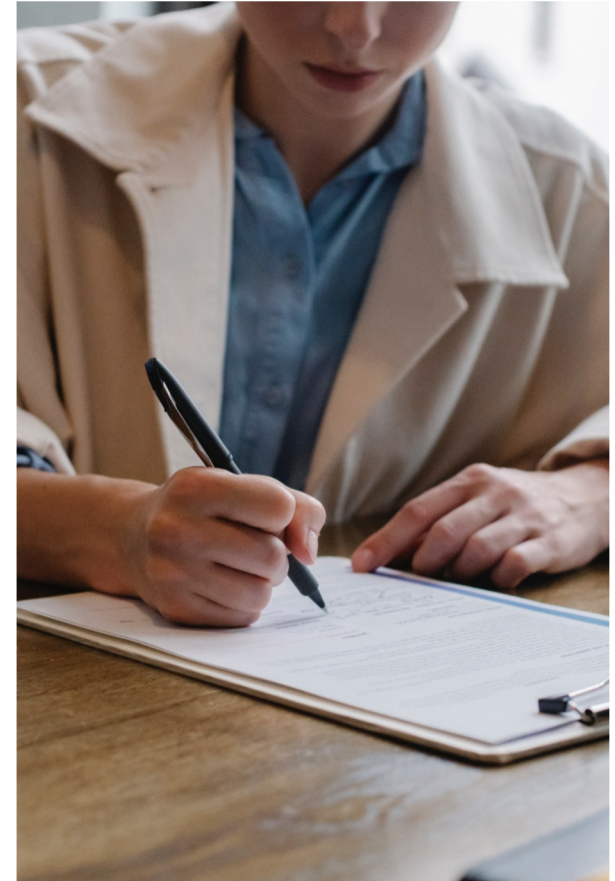 - Mobile floating data
  - Tracking apps

Photo by Sora Shimazaki from Pexels

Our Goal

Find out if (or how) tweet sentiment can be used as a data source for urban planning

J﹀U

Our Approach

- Use social media posts (tweets) for evaluating how citizens interact with their city

- Continuous stream of "live" data

- Insight into sentiment of citizens

Photo by Solen Feyissa from Pexels

JⴢU

Research Questions

- RQ1: Are there differences in Tweet sentiment in different cities?

- RQ2: Are there differences in Tweet sentiment at certain POIs or (POI types) within a city

- RQ3: Are there differences in Tweet sentiment over time? Are these more distinct at certain POIs?

JⴸU

Dataset

- Geolocated Tweets from Twitter API

- Worldwide Subset of total tweets

- Focus on 10 selected metropolitan regions

- September 1 – September 17 2012

JᴗU

Dataset

Metropolitan Area	# EN Tweets	% EN Tweets	Population
New York	303.3 K	80.9%	9.4 M
Los Angeles	225.4 K	76.9%	5.0 M
London	198.2 K	78.2%	7.7 M
Manila	145.6 K	55.5%	3.1 M
Chicago	138.7 K	79.5%	3.7 M
Jakarta	87.7 K	12.3%	9.7 M
Tokyo	27.4 K	15.0%	13.8 M
Paris	27.3 K	11.8%	7.5 M
Mexico City	24.0 K	13.3%	10.8 M
Sydney	21.0 K	81.6%	2.7 M
Lagos	20.7 K	66.2%	7.1 M
Melbourne	19.6 K	80.8%	1.9 M
Istanbul	17.7 K	5.9%	9.9 M
Sao Paolo	16.5 K	7.5%	12.5 M
Johannesburg	12.0 K	77.4%	3.9 M

JƱU

Processing – Overview
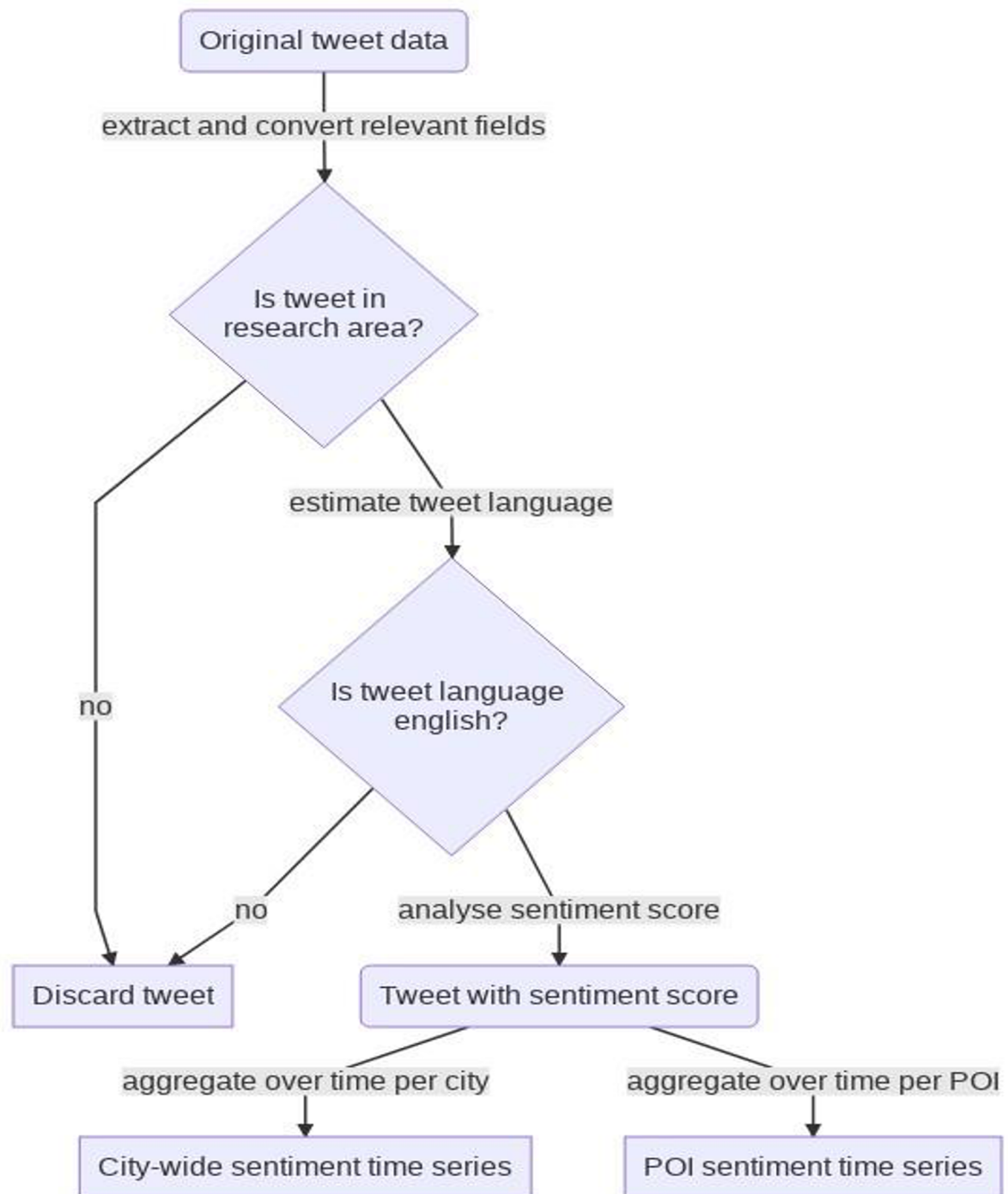
Tweet Pre-Processing

before

	in_reply_to_status_id_str	id_str	text	in_reply_to_screen_name	in_reply_to_user_id_str	favorited	source	entities	truncated	created_at	...	geo	retweet_count	retweeted	coordinates	user	id	in_reply_to_status_id
0	247583648478535680	247595834722631680	@ShabyAdebat oui j'y suis déjà la	ShabyAdebat	454690071	False	Twitter for BlackBerry®	{'hashtags': [], 'user_mentions': [{'id_str': '454690071', 'indices': [0, 12], 'screen_name': 'ShabyAdebat', 'name': 'Shaby', 'id': 454690071}], '...	False	Mon Sep 17 07:20:21 +0000 2012	...	{'type': 'Point', 'coordinates': [48.62580937, 2.56456936]}	0.0	False	{'type': 'Point', 'coordinates': [2.56456936, 48.62580937]}	{'id_str': '563884385', 'follow_request_sent': None, 'default_profile_image': False, 'profile_use_background_image': True, 'friends_count': 41, 'p...	2.475958e+17	2.475836e+17
1	247595108483092480	247595835108511744	@Norshafiqashuib a a ku ku, aku kaki sakitttt! haaa, kaki sakit.. haha.	Norshafiqashuib	543294831	False	web	{'hashtags': [], 'user_mentions': [{'id_str': '543294831', 'indices': [0, 16], 'screen_name': 'Norshafiqashuib', 'name': '30 September :)', 'id': ...	False	Mon Sep 17 07:20:21 +0000 2012	...	None	0.0	False	None	{'id_str': '710651022', 'follow_request_sent': None, 'profile_use_background_image': True, 'friends_count': 140, '...	2.475958e+17	2.475951e+17
2	None	247595835330809856	WhatsApp @alyafatysa	None	None	False	Twitter for iPhone	{'hashtags': [], 'user_mentions': [{'id_str': '323637488', 'indices': [9, 20], 'screen_name': 'alyafatysa', 'name': 'Fatysa Ramlan', 'id': 3236374...	False	Mon Sep 17 07:20:21 +0000 2012	...	{'type': 'Point', 'coordinates': [3.09893107, 101.55374855]}	0.0	False	{'type': 'Point', 'coordinates': [101.55374855, 3.09893107]}	{'id_str': '579451411', 'follow_request_sent': None, 'default_profile_image': False, 'profile_use_background_image': True, 'friends_count': 44, 'p...	2.475958e+17	NaN
3	247594287238369280	247595835284676608	@08misaki18 塾頑張って下さいねぇ(*^^*)また お邪魔させていただきまぁす♪	08misaki18	462837743	False	Twitter for Android	{'hashtags': [], 'user_mentions': [{'id_str': '462837743', 'indices': [0, 11], 'screen_name': '08misaki18', 'name': '美咲', 'id': 462837743}], 'urls...	False	Mon Sep 17 07:20:21 +0000 2012	...	{'type': 'Point', 'coordinates': [34.466747, 135.3721579]}	0.0	False	{'type': 'Point', 'coordinates': [135.3721579, 34.466747]}	{'id_str': '558801706', 'follow_request_sent': None, 'default_profile_image': False, 'profile_use_background_image': True, 'friends_count': 227, '...	2.475958e+17	2.475943e+17

after

	id	city_id	created_at	text	geom	sentiment	cemetery	park	transport	minute	hour	dow	leveled_sentiment
0	103960	3746	2012-10-07 13:27:05	there hasn't been much to be positive about in...	POINT (-73.84604 40.73867)	-0.697527	None	True	None	27	13	6	-0.738166
1	104382	3746	2012-10-07 13:29:38	its 930am on a sunday.. idk what cause people ...	POINT (-73.56045 40.73845)	-0.924843	None	True	None	29	13	6	-0.965482
2	104003	3746	2012-10-07 13:27:22	I'm fucking late for my orientation	POINT (-73.83158 40.68454)	-0.972891	None	None	True	27	13	6	-1.013530
3	104142	3746	2012-10-07 13:28:16	@user I love drea dematteo and now that I I kn...	POINT (-73.92970 40.75568)	0.986070	None	None	True	28	13	6	0.945431
4	104447	3746	2012-10-07 13:30:06	@user ran in to @user the other day ! lets gra...	POINT (-73.98120 40.72773)	0.815802	None	None	None	30	13	6	0.775163

JⱯU

Sentiment Analysis

- RoBERTa pre-trained Model
  - (provided by [NLP group at University of Cardiff](#))
  - Requires pre-processing (links, users, …)

- Result:

```
positive 0.8466
neutral 0.1458
negative 0.0076
```

Sentiment Analysis
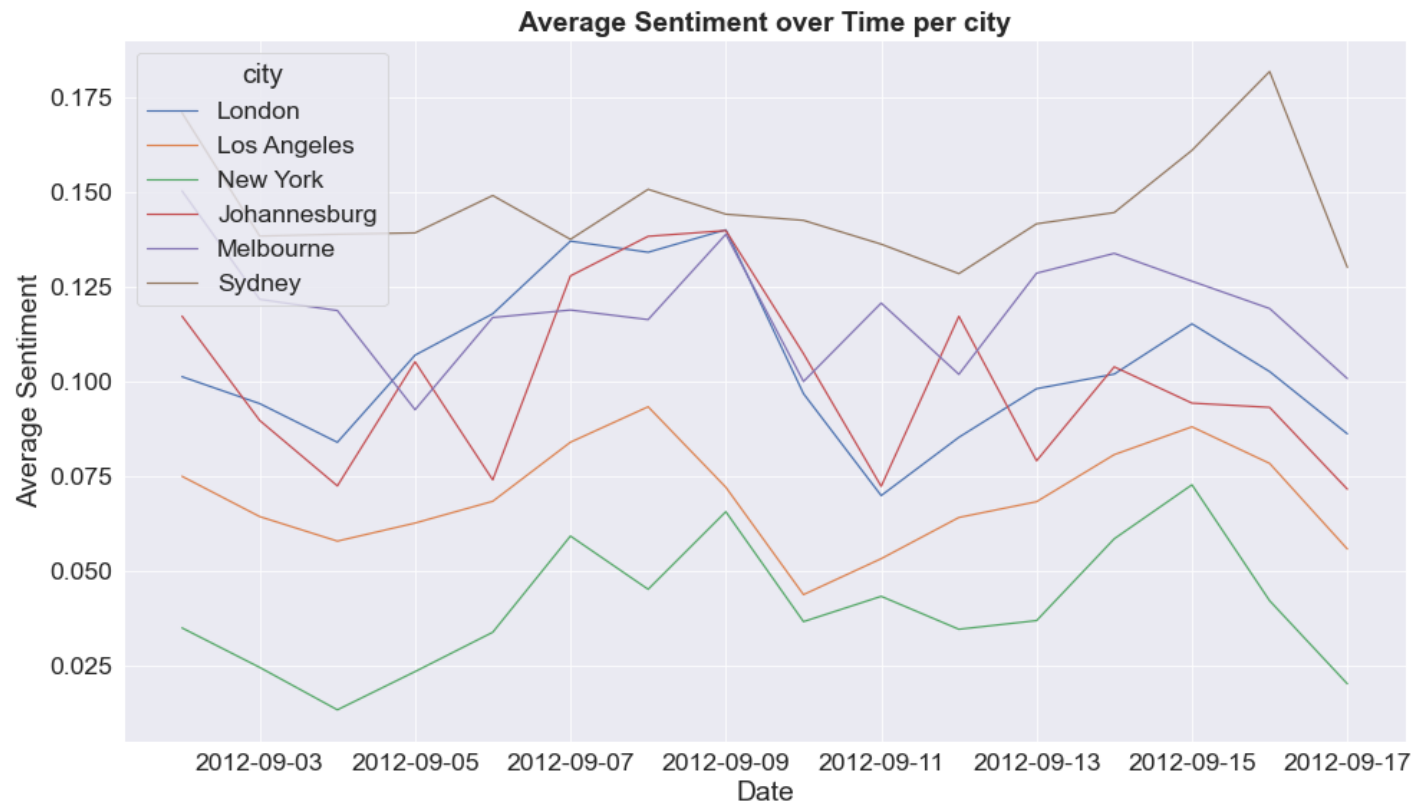
- Calculate sentiment-score

$$s(t_i) = \begin{cases} 0, & \text{if } p(t_i = neu) > p(t_i = pos) \wedge \\ & \qquad p(t_i = neu) > p(t_i = neg) \\ p(t_i = pos) - p(t_i = neg), & \text{otherwise} \end{cases}$$

- Calculate adjusted sentiment for POI comparison

- Aggregate based on city and POI influence area

Results

RQ1: Are There Differences in Tweet Sentiment in Different Cities?

Average Sentiment over Time per city

Adding POI Information

- Data from OpenStreetMap (OSM)
- Focus on cemeteries, transit nodes (e.g. railway or bus stations) and parks
- Can be "mapped" in OSM network differently
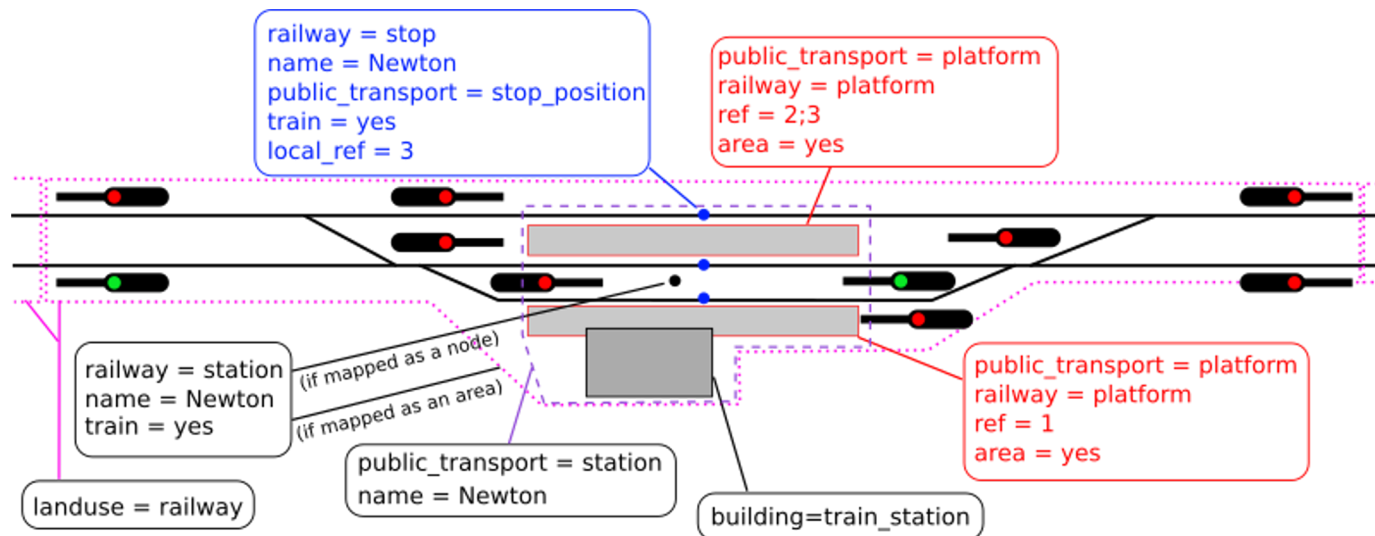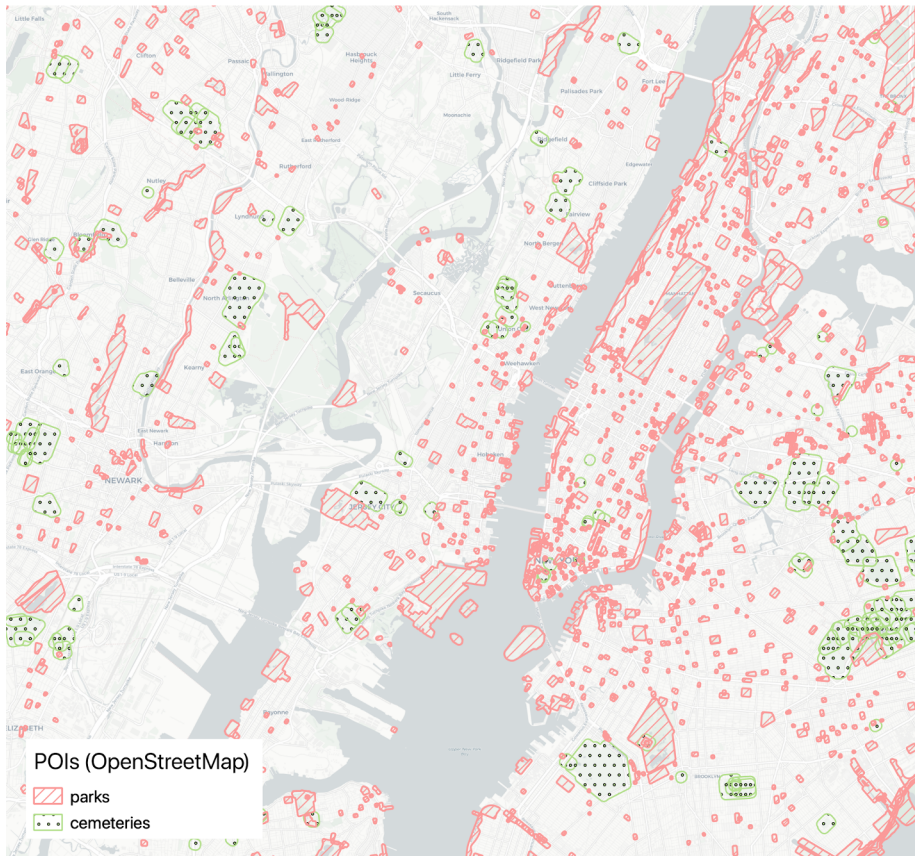 - point / line / polygon
  - Data as "tags"

railway = stop
name = Newton
public_transport = stop_position
train = yes
local_ref = 3

public_transport = platform
railway = platform
ref = 2;3
area = yes

railway = station (if mapped as a node)
name = Newton (if mapped as an area)
train = yes

public_transport = platform
railway = platform
ref = 1
area = yes

landuse = railway

public_transport = station
name = Newton

building=train_station

Illustration from wiki.openstreetmap.org

Adding POI Information

POIs (OpenStreetMap)
- parks
- cemeteries
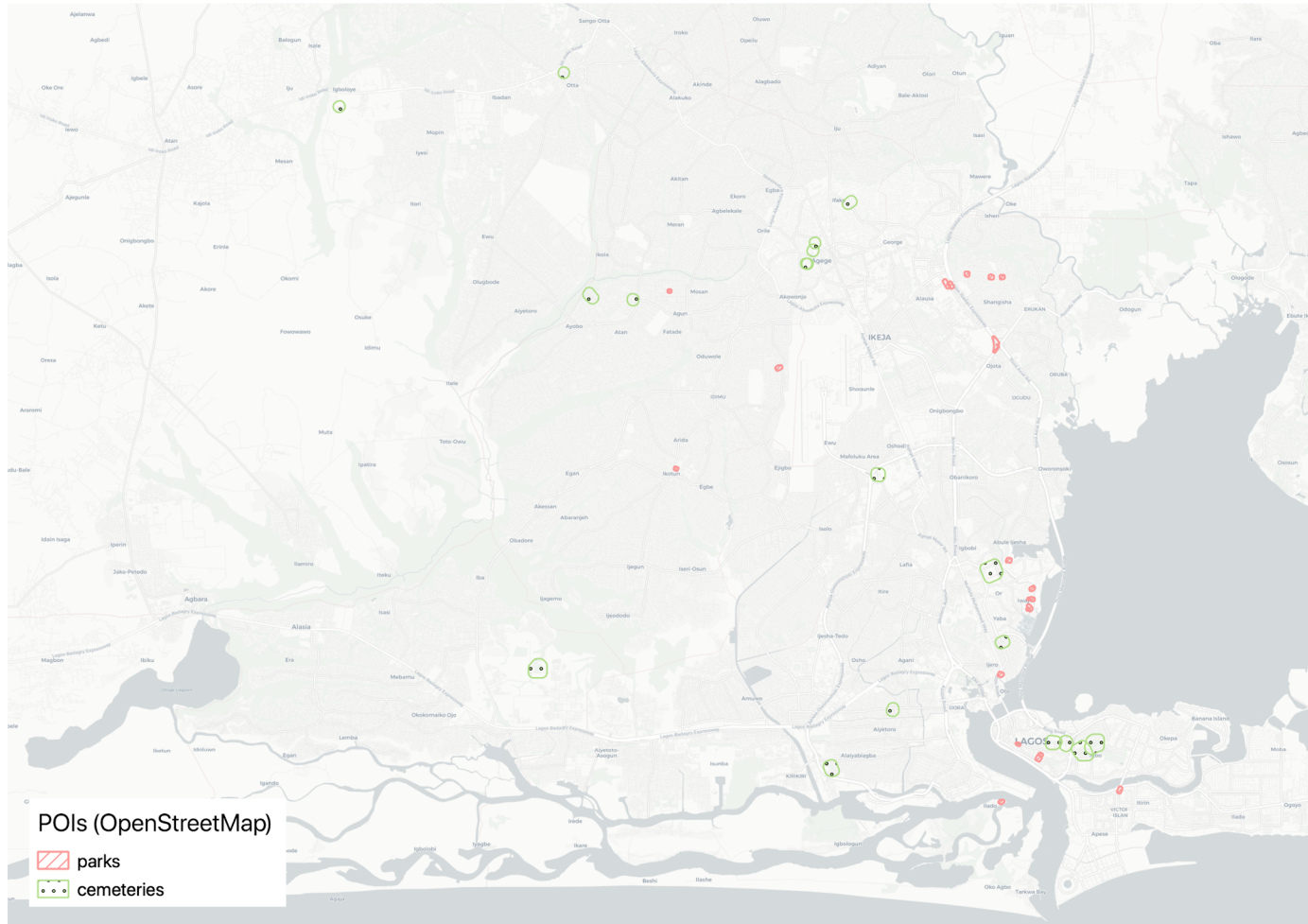
New York Metropolitan Area

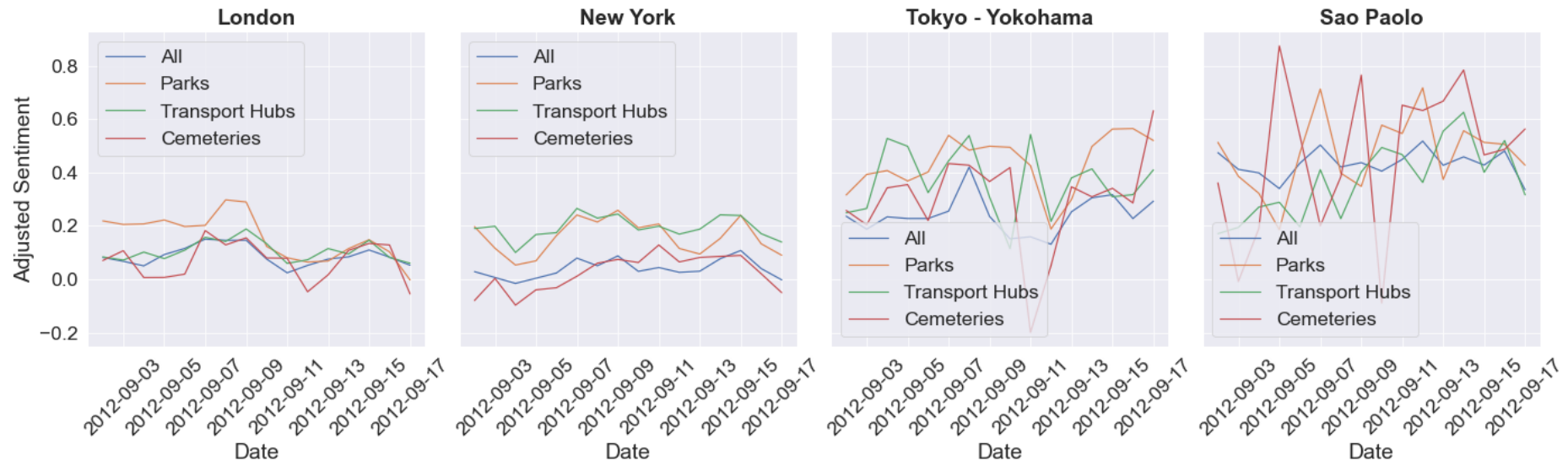

POIs (OpenStreetMap)
- parks
- cemeteries
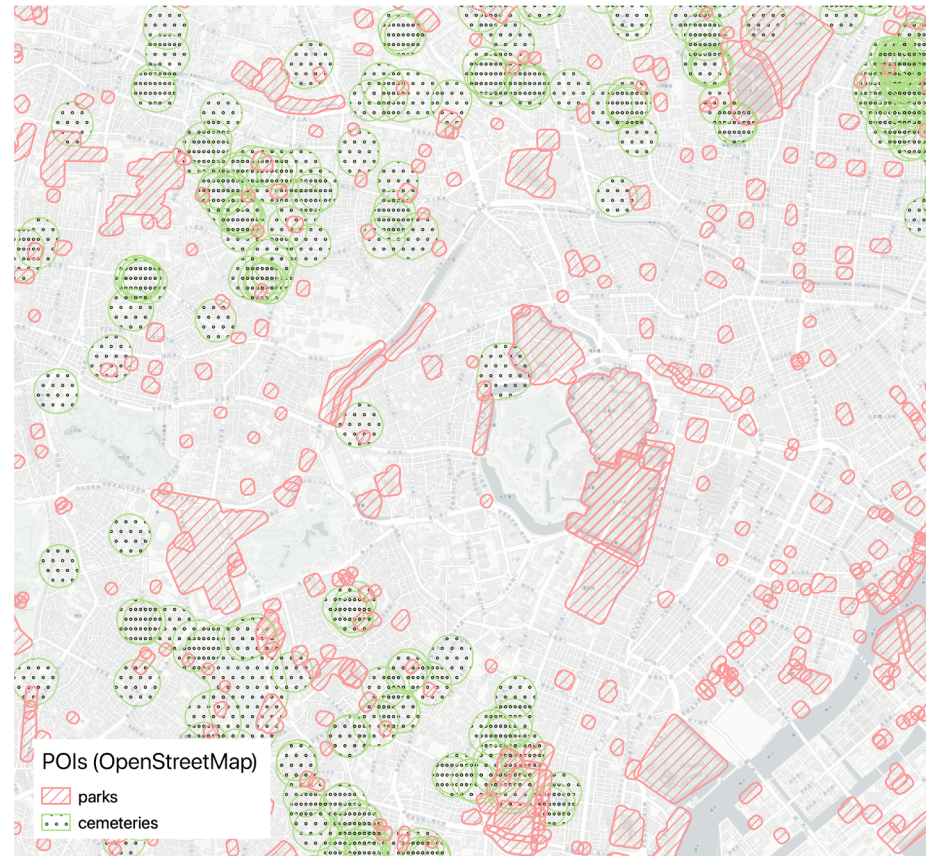
Tokyo-Yokohama Metropolitan Area
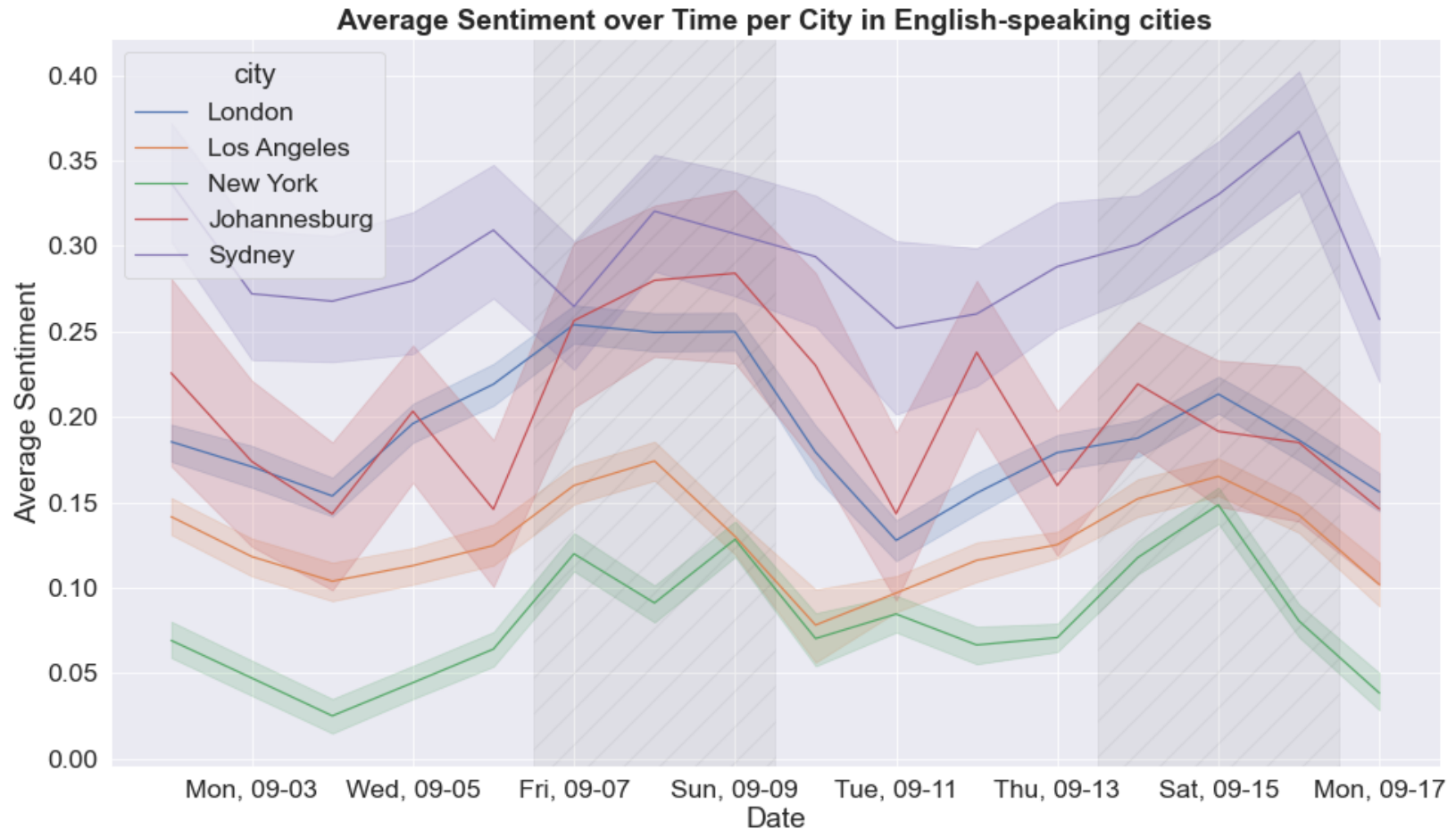
Adding POI Information

Lagos Metropolitan Area

RQ2: Are There Differences in Tweet Sentiment at Certain POIs in a City

RQ3: Are There Differences in Tweet Sentiment Over Time?

Average Sentiment over Time per City in English-speaking cities

Other Takeaways

- Quality and quantity of OpenStreetMap data differs greatly between regions

- Non-English-speaking cities show weaker relation between POI and sentiment

- Filtering out tourists' Tweets needed for urban planning application

JↃU

Future Work

- Analyzing change of general sentiment and sentiment at POIs during / after CoViD-19

- Finding clusters of high / low sentiment in a city

- Distinguish between tweets by tourists and city inhabitants

JⴲU

Thank you!

Chris Stelzmüller
chris@triply.at
Twitter: @chrisstelz