Deriving Spatio-Temporal Geographies from mobile GPS data: A data-driven approach

Living Boundaries project

1. Nombuyiselo Murage – MSc. Geographic Data Science.
2. Dr. Alessia Calafiore – University of Liverpool
3. Franck Lenoir and Irem Ozen – Tamoco UK Limited

Problem Statement

One of the initial questions posed to the researcher, was, whether it is possible to quantify and model the collective experiences of city dwellers and after that, determine regions that have a shared common experience.

Methodology

Tamoco Limited UK consists of **1,195,952,837** **(approximately 1.2 billion data points)** of mobile phone GPS traces that have been generated by **6,761,537 unique users** in New York City.

Outlier Detection

Complimentary Tool for Erroneous Data Filtering

H3 Trajectory Compression

Stops Detection & Compressing the Data

Home – Work - Leisure Detection

Filtering for Leisure Stops & Removing Sensitive Locations

Community Detection

Identifying communities

Temporal Profiles

Understanding the impact of temporal signatures

trajectory (x_i, y_i, t_i)

trajectory (x_j, y_j, t_j)

trajectory (x_k, y_k, t_k)

1 2 3

(a) (b) (c)

GOAL OF DATA PRE-PROCESSING

Low dimensional representation of trajectories. To represent a user's trajectory as a sequence of grid cells
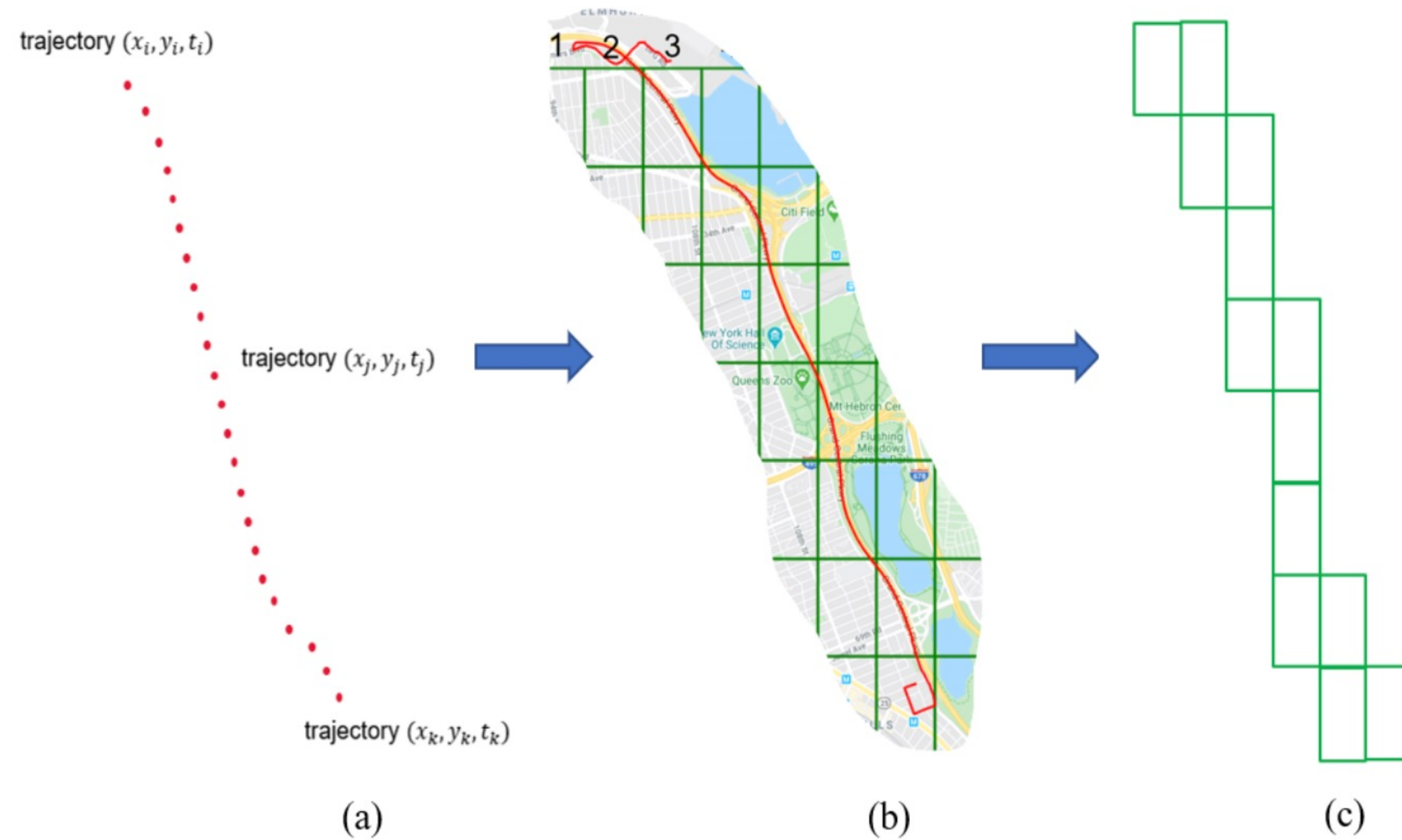
Uber H3 Cell

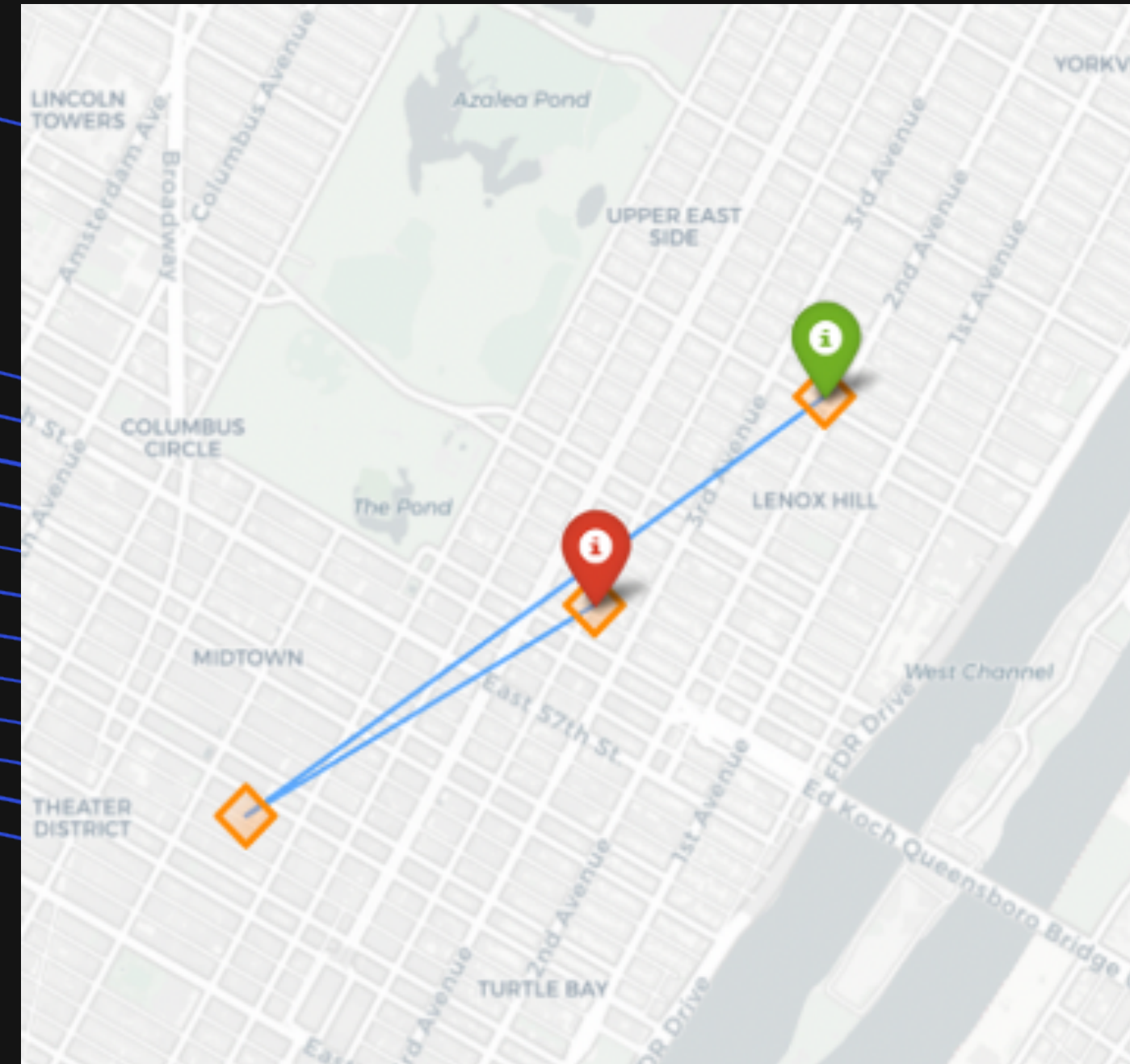## Geoprivacy – Uber Resolution Level 9

Data Pre-Processing

(a) A single user's trajectory represented as a sequence of H3 hex cells before data pre-processing.

(b) stops detected along the same user's trajectory denoted by the center coordinates of the H3 cells identified as stops.

Highlights of the H3 Trajectory Compression

device_id	date	sdk_ts	hex_id	geo_point	
2a12025c-815c-48ee-a0c5-0bee05e96aea	10/7/19	2019-10-07 06:23:07 UTC	892a100892fffff	POINT(-73.95864807193321 40.7691614091953)	User 1
2a12025c-815c-48ee-a0c5-0bee05e96aea	10/7/19	2019-10-07 06:24:04 UTC	892a100892fffff	POINT(-73.95864807193321 40.7691614091953)	
2a12025c-815c-48ee-a0c5-0bee05e96aea	10/7/19	2019-10-07 06:58:25 UTC	892a100892fffff	POINT(-73.95864807193321 40.7691614091953)	
2a12025c-815c-48ee-a0c5-0bee05e96aea	10/7/19	2019-10-07 06:58:26 UTC	892a100892fffff	POINT(-73.95864807193321 40.7691614091953)	Stop 1
2a12025c-815c-48ee-a0c5-0bee05e96aea	10/7/19	2019-10-07 06:58:30 UTC	892a100892fffff	POINT(-73.95864807193321 40.7691614091953)	
2a12025c-815c-48ee-a0c5-0bee05e96aea	10/7/19	2019-10-07 06:59:37 UTC	892a100892fffff	POINT(-73.95864807193321 40.7691614091953)	
2a12025c-815c-48ee-a0c5-0bee05e96aea	10/7/19	2019-10-07 07:00:13 UTC	892a100892fffff	POINT(-73.95864807193321 40.7691614091953)	
2a12025c-815c-48ee-a0c5-0bee05e96aea	10/7/19	2019-10-07 07:00:44 UTC	892a100892fffff	POINT(-73.95864807193321 40.7691614091953)	End
2a12025c-815c-48ee-a0c5-0bee05e96aea	10/7/19	2019-10-07 07:00:59 UTC	892a100d697fffff	POINT(-73.9606847585308 40.7664624719519)	Start
2a12025c-815c-48ee-a0c5-0bee05e96aea	10/7/19	2019-10-07 07:01:31 UTC	892a100d697fffff	POINT(-73.9606847585308 40.7664624719519)	
2a12025c-815c-48ee-a0c5-0bee05e96aea	10/7/19	2019-10-07 07:01:40 UTC	892a100d697fffff	POINT(-73.9606847585308 40.7664624719519)	
2a12025c-815c-48ee-a0c5-0bee05e96aea	10/7/19	2019-10-07 07:03:31 UTC	892a100d693fffff	POINT(-73.9629257194079 40.7690467484428)	
2a12025c-815c-48ee-a0c5-0bee05e96aea	10/7/19	2019-10-07 07:03:37 UTC	892a100d693fffff	POINT(-73.9629257194079 40.7690467484428)	
2a12025c-815c-48ee-a0c5-0bee05e96aea	10/7/19	2019-10-07 07:05:45 UTC	892a100d68bfffff	POINT(-73.9692395588398 40.7662328404129)	Weird Outlier
2a12025c-815c-48ee-a0c5-0bee05e96aea	10/7/19	2019-10-07 07:05:51 UTC	892a100d68fffff	POINT(-73.9669983736748 40.7636487730393)	
2a12025c-815c-48ee-a0c5-0bee05e96aea	10/7/19	2019-10-07 07:05:56 UTC	892a100d68fffff	POINT(-73.9669983736748 40.7636487730393)	
2a12025c-815c-48ee-a0c5-0bee05e96aea	10/7/19	2019-10-07 07:06:01 UTC	892a100d613fffff	POINT(-73.971275526503 40.763533802293)	
2a12025c-815c-48ee-a0c5-0bee05e96aea	10/7/19	2019-10-07 07:09:20 UTC	892a100d677fffff	POINT(-73.977382044752 40.7554369941077)	
2a12025c-815c-48ee-a0c5-0bee05e96aea	10/7/19	2019-10-07 07:09:34 UTC	892a100d63bfffff	POINT(-73.97331056239359 40.7555521924493)	
2a12025c-815c-48ee-a0c5-0bee05e96aea	10/7/19	2019-10-07 07:09:41 UTC	892a100d63bfffff	POINT(-73.97331056239359 40.7555521924493)	
2a12025c-815c-48ee-a0c5-0bee05e96aea	10/7/19	2019-10-07 07:09:49 UTC	892a100d63bfffff	POINT(-73.97331056239359 40.7555521924493)	
2a12025c-815c-48ee-a0c5-0bee05e96aea	10/7/19	2019-10-07 08:15:22 UTC	892a100d677fffff	POINT(-73.977382044752 40.7554369941077)	
2a12025c-815c-48ee-a0c5-0bee05e96aea	10/7/19	2019-10-07 08:15:30 UTC	892a100d677fffff	POINT(-73.977382044752 40.7554369941077)	
2a12025c-815c-48ee-a0c5-0bee05e96aea	10/7/19	2019-10-07 08:23:33 UTC	892a100d677fffff	POINT(-73.977382044752 40.7554369941077)	
2a12025c-815c-48ee-a0c5-0bee05e96aea	10/7/19	2019-10-07 08:27:54 UTC	892a100d677fffff	POINT(-73.977382044752 40.7554369941077)	
2a12025c-815c-48ee-a0c5-0bee05e96aea	10/7/19	2019-10-07 08:30:59 UTC	892a100d677fffff	POINT(-73.977382044752 40.7554369941077)	
2a12025c-815c-48ee-a0c5-0bee05e96aea	10/7/19	2019-10-07 08:46:14 UTC	892a100d677fffff	POINT(-73.977382044752 40.7554369941077)	
2a12025c-815c-48ee-a0c5-0bee05e96aea	10/7/19	2019-10-07 10:28:57 UTC	892a100d677fffff	POINT(-73.977382044752 40.7554369941077)	
2a12025c-815c-48ee-a0c5-0bee05e96aea	10/7/19	2019-10-07 10:56:38 UTC	892a100d677fffff	POINT(-73.977382044752 40.7554369941077)	
2a12025c-815c-48ee-a0c5-0bee05e96aea	10/7/19	2019-10-07 12:00:36 UTC	892a100d677fffff	POINT(-73.977382044752 40.7554369941077)	Stop 2
2a12025c-815c-48ee-a0c5-0bee05e96aea	10/7/19	2019-10-07 12:36:02 UTC	892a100d677fffff	POINT(-73.977382044752 40.7554369941077)	

Highlights of the H3 Trajectory Compression

Achieved 92.7% compression on Average across all 3 months

Month	Initial Total	After Filtering Accuracy	Outlier Detection	H3 Compression	% Reduction
October	427,116,689	393,446,648	387,471,916	31,542,980	91.9 %
November	411,301,228	385,480,446	378,553,555	25,814,050	93.2 %
December	357,534,920	338,570,073	331,538,063	23,019,261	93.1 %

Spatially Augmented Community Detection

- SPATIAL INTER DEPENDENCY: can be defined as the connection between two places based on a common set of visitors to those places over a period of time (temporal frequency)
- If a user visits a particular venue, there is a relationship (custom weight).
- There is a relationship between two venues, if a common set of users visit these two locations.
- If a subset of these users visit another venue, then we can say that there is a relationship between the new venue and the previous venues i.e there is some similarity that is inferred due to the fact that it is being visited by a similar group of users (semantic proximity!).

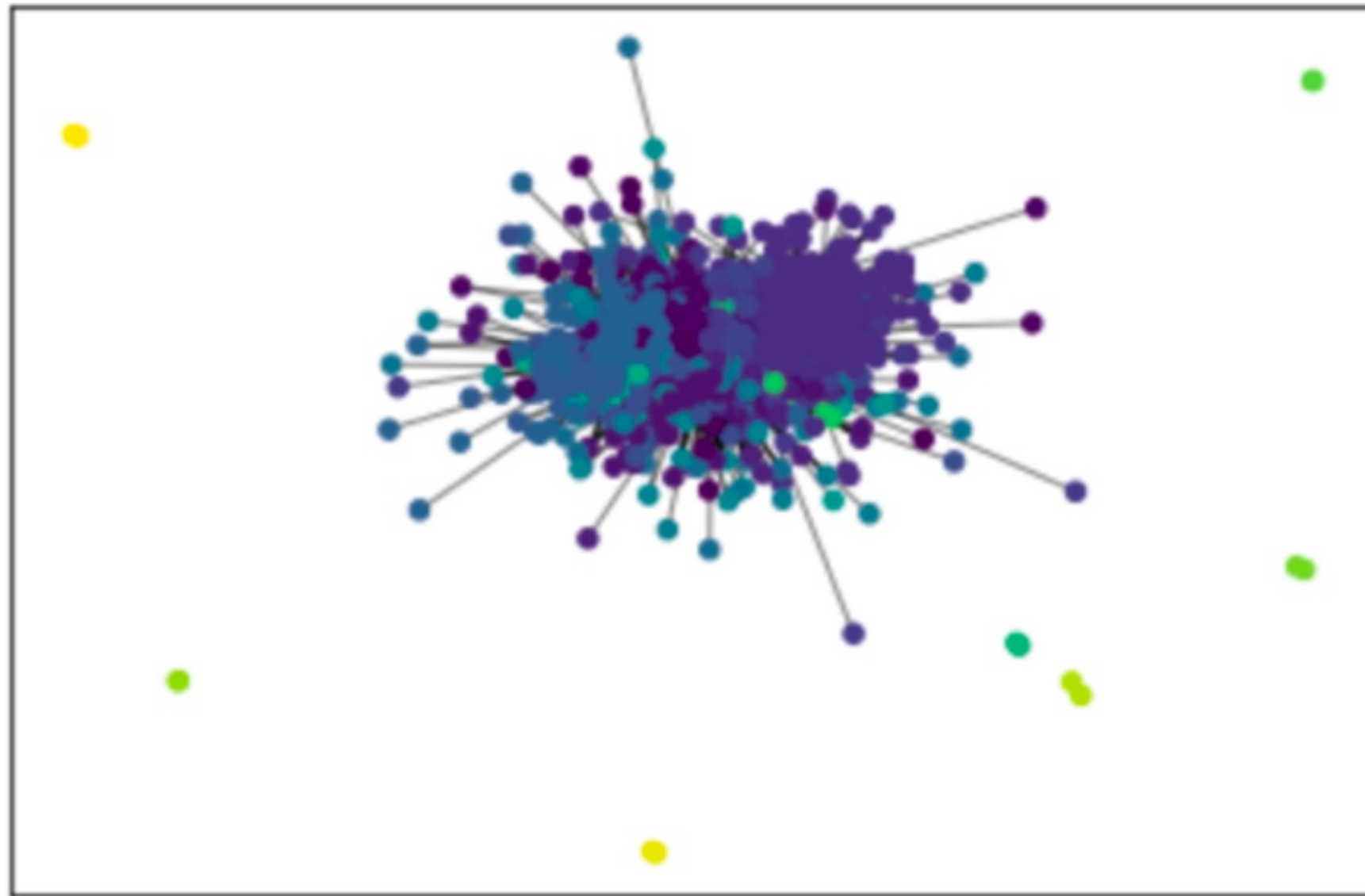| Row	device_id	hex_id	visits_count	avg_duration	weight
1	de0cf9e2-0253-404e-99ab-be62a31e49d2	892a1077457ffff	17	188.82352941176472	0.757552946580887
2	1d3f3563-af37-45cd-97c6-fbf79aa98789	892a100ab17ffff	18	311.5555555555556	0.809900605896093
3	34105f95-6645-4437-87ad-72725ff81c62	892a103826fffff	26	265.19230769230774	0.82928526276309267
4	7e83e545-3133-4389-836f-86f829d3c4aa	892a1074693ffff	20	112.25000000000001	0.72400404208195657
5	d4f34074-6ead-4794-bdf0-5efdd36ec69c	892a100a8afffff	22	402.00000000000006	0.8526425457431136
6	448b4c65-424f-480e-9a16-e0911a72f7c6	892a100f3afffff	32	325.15625	0.86789368620715768

Spatially Augmented Community Detection
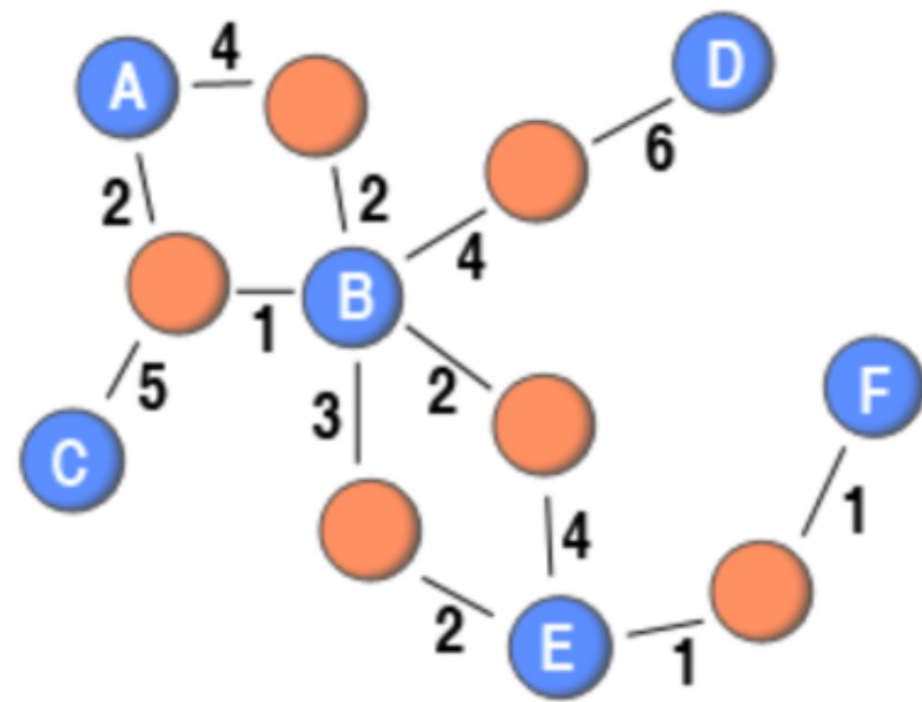
Figure 4.12: Visual Image to Illustrate our user-to-place Bipartite Graph

Figure 4.13: Visual Illustration of the Projection Process

- Individual user-to-place interactions (activity patterns from Patterson & Farber (2015)).
- Spatial proximity (from Tobler's first law, "nearer objects are more related than distant objects" ?).
- Temporal signatures, based on the filtering the data by the various temporal frequency, described above.

Spatially Augmented Community Detection

- Individual user-to-place interactions (activity patterns from Patterson & Farber (2015)).
- Spatial proximity (from Tobler's first law, "nearer objects are more related than distant objects" ?).
- Temporal signatures, based on the filtering the data by the various temporal frequency, described above.

Demo