Enhancing Markov-chain spatial simulation with web mapping technology

Chun Fung Kwok (Jackson)

St. Vincent's Institute of Medical Research (SVI)

2021/07/30 @ ECSSC 2021

Outline

1. Motivation

2. Working with web mapping data in R

- The data
- Basic queries and visualisation
- Converting map data into a graph

3. Markov Chain on graph

- The model
- Enhancement with web mapping

(Slides are available at `https://github.com/kcf-jackson/ecssc-2021-talk/`.)

Motivation

A few simple questions...

"How many cafes / restaurants are there in Melbourne CBD?"

"What does the flow between suburbs look like over a typical day?"

"If a road is blocked, how much extra time does it cost?"

"Given a residential address, how good is its access to shops, supermarkets, restaurants, hospitals, etc.?"

To answer this type of questions, we need:

1. web mapping data (OpenStreetMap data, loaded into R with `osmextract`)
2. the capacity to perform spatial simulation and
3. manipulate the map interactively to display information (`sketch` and `leaflet.js`)

Working with web mapping data in R

The data

```r
library(osmextract)
library(sf)
library(dplyr)

file <- "./data/melbourne.osm.pbf"      # Load data
feature_counts <- st_layers(file)       # Get a summary of what's available
```

```
> feature_counts
Driver: OSM
Available layers:
        layer_name       geometry_type features fields
1           points               Point       NA     10
2            lines         Line String       NA      9
3 multilinestrings   Multi Line String       NA      4
4    multipolygons       Multi Polygon       NA     25
5  other_relations Geometry Collection       NA      4
```

Working with web mapping data in R

The data

```
> melb_points <- oe_read(file, "points")  # Import the features
> names(melb_points)
 [1] "osm_id"     "name"       "barrier"    "highway"
 [5] "ref"        "address"    "is_in"      "place"
 [9] "man_made"   "other_tags" "geometry"
```

```
> head(melb_points[, c("osm_id", "other_tags", "geometry")])
Simple feature collection with 6 features and 2 fields
Geometry type: POINT
Dimension:     XY
Bounding box:  xmin: 145.1294 ymin: -37.91224 xmax: 145.1368 ymax: -37.90771
Geodetic CRS:  WGS 84
  osm_id                                        other_tags                    geometry
1 579287                          "traffic_calming"=>"hump" POINT (145.1352 -37.90783)
2 579296                          "traffic_calming"=>"hump" POINT (145.1294 -37.90771)
3 579301                          "traffic_calming"=>"hump"   POINT (145.13 -37.90845)
4 579306                          "traffic_calming"=>"hump" POINT (145.1368 -37.90978)
5 579314                                               <NA> POINT (145.1362 -37.91224)
6 579338 "crossing"=>"uncontrolled","crossing_ref"=>"zebra" POINT (145.1365 -37.91158)
```

Working with web mapping data in R

Basic queries

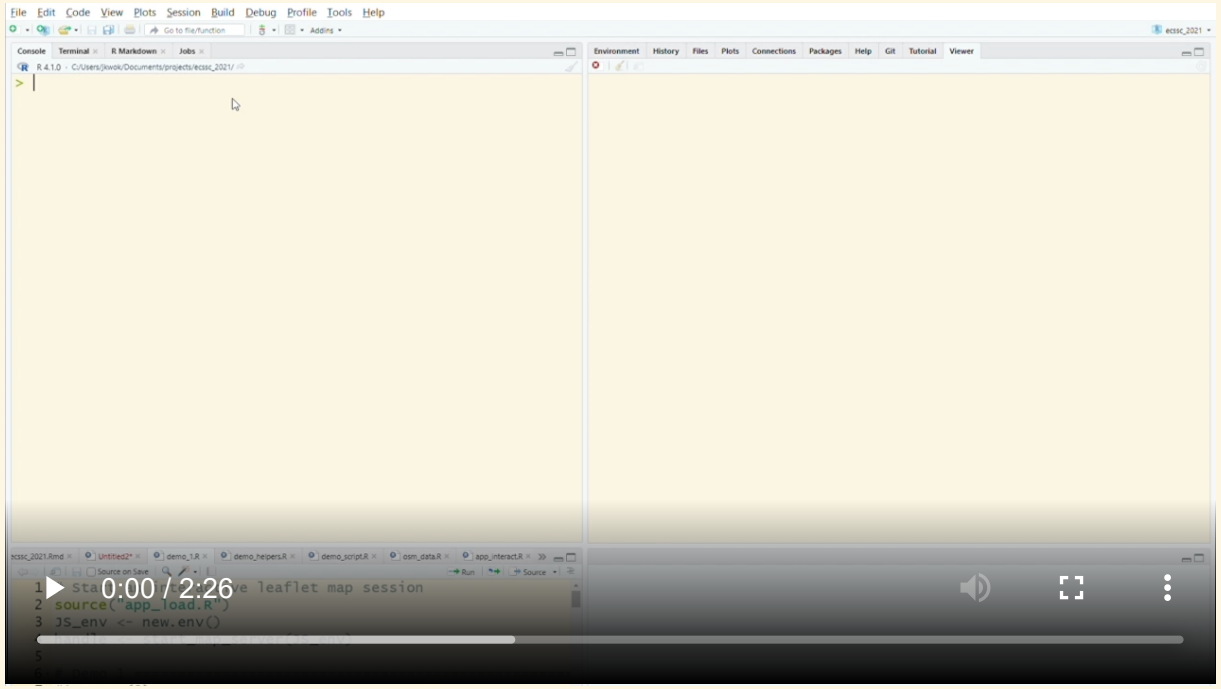
```
# Number of cafes
> sum(grepl("cafe", melb_points$other_tags))
[1] 1918
```

```
# Number of restaurants in Melbourne
> is_restaurant <- grepl("restaurant", melb_points$other_tags)
> sum(is_restaurant)
[1] 2360
```

```
# Number of restaurants in Melbourne CBD
> melb_mpolygons <- oe_read(file, "multipolygons")
> melb_CBD_boundary <- melb_mpolygons |>
+   filter(name == "City of Melbourne") |>
+   get_points_matrix()
>
> CBD_restaurants <- melb_points[is_restaurant, ] |>
+   filter(in_bound(geometry, melb_CBD_boundary))
> nrow(CBD_restaurants)
[1] 605
```

Working with web mapping data in R

Basic visualisation (with `sketch` and `leaflet.js`)



Working with web mapping data in R

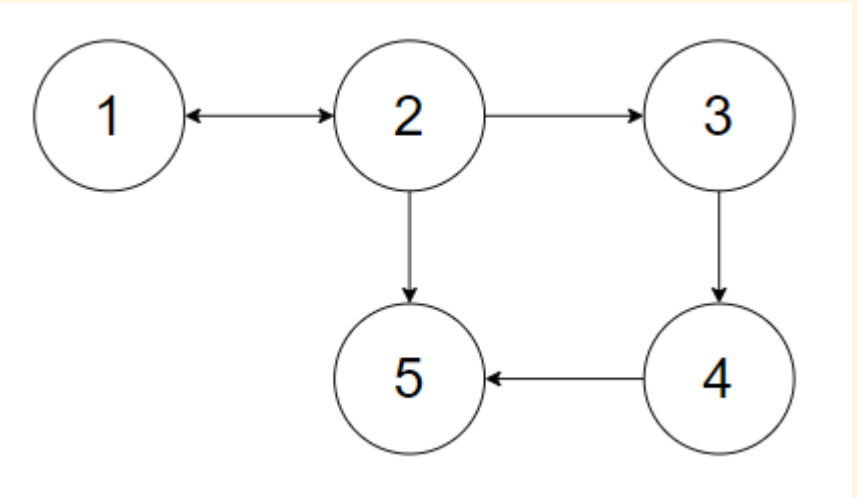
Converting map data into a graph

To simulate movement on a map or compute geodesic distance between two points, a graph is needed.

A graph encodes

- a node-list: a list of "meaningful" points on the map
  RHS: {1,2,3,4,5}
- a edge-list: a list of pairs of nodes, indicating the link between the two nodes.
  RHS: {(1,2), (2,1), (2,3), (2,5), (3,4), (4,5)}

An example of a graph



Working with web mapping data in R

Converting map data into a graph

Suburb graph

A suburb is represented by its boundary (i.e. a set of coordinates). To construct a suburb graph,

1. give each suburb an unique node ID and add to the node-list;
2. for any two suburbs that share some boundary points, add an edge between them to the edge-list.

Working with web mapping data in R

Converting map data into a graph

Street graph

A street is represented by a sequence of coordinates. To construct a street graph, for each street,

1. give each coordinate an unique node ID and add to the node-list
   - points with the same coordinate should have the same ID
   - the node-list should have contain duplicate entries
2. add an edge for every two consecutive nodes
   - if the street is two-way, then also add the edge for the reverse direction

Once the graph is created, one can use existing algorithms to find the route between any two nodes, e.g. the A* path-finding algorithm and the multi-level Dijkstra (MLD) algorithm.

(Alternatively, the `osrm` package handles both the graph building and path-finding for you.)

The model: Markov chain on graphs

Let $S = \{1, 2, , \ldots, n\}$ be the index set of the n nodes ("suburbs") on a graph.

Suppose there are m IID agents travelling on the graph, and

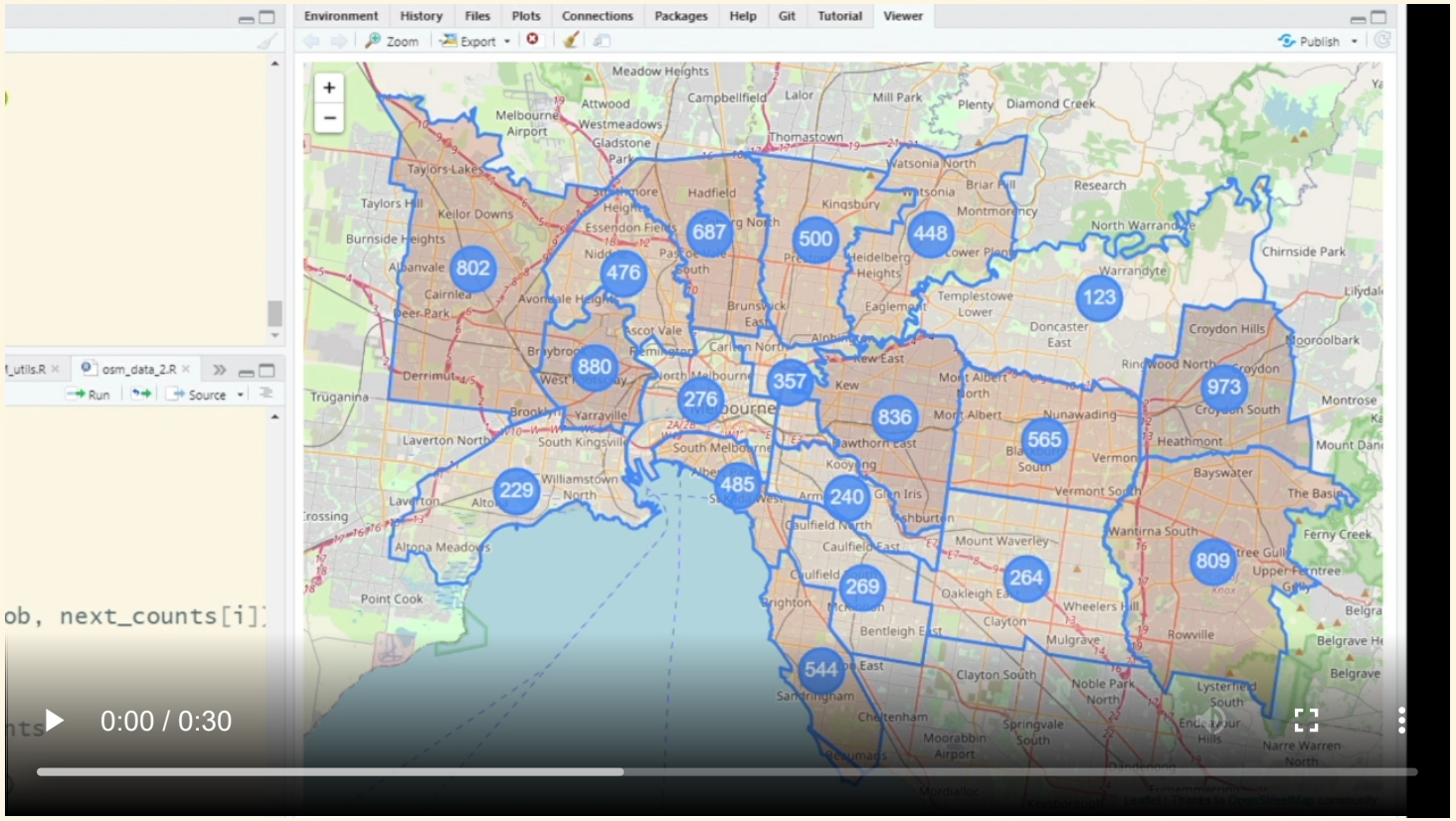
- let $A_k(t)$ represent the k-th agent and record where the agent is located at discrete time $t \in \mathbb{N}$.
- The transition is modelled using a Markov chain with transition probability $P\left(A_k(t+1) = s | A_k(t) = s'\right)$.

For any node / location $s \in S$,

- count the number of agents present at time t, $X_s(t) = \sum_{k=1}^{m} 1_{A_k(t)=s}$.

We are interested in the dynamics of $X_s(t)$ for $s \in S, t \in \mathbb{N}$

Demo


Environment
History
Files
Plots
Connections
Packages
Help
Git
Tutorial
Viewer
Zoom
Export
Publish
Run
Source
0:00 / 0:30

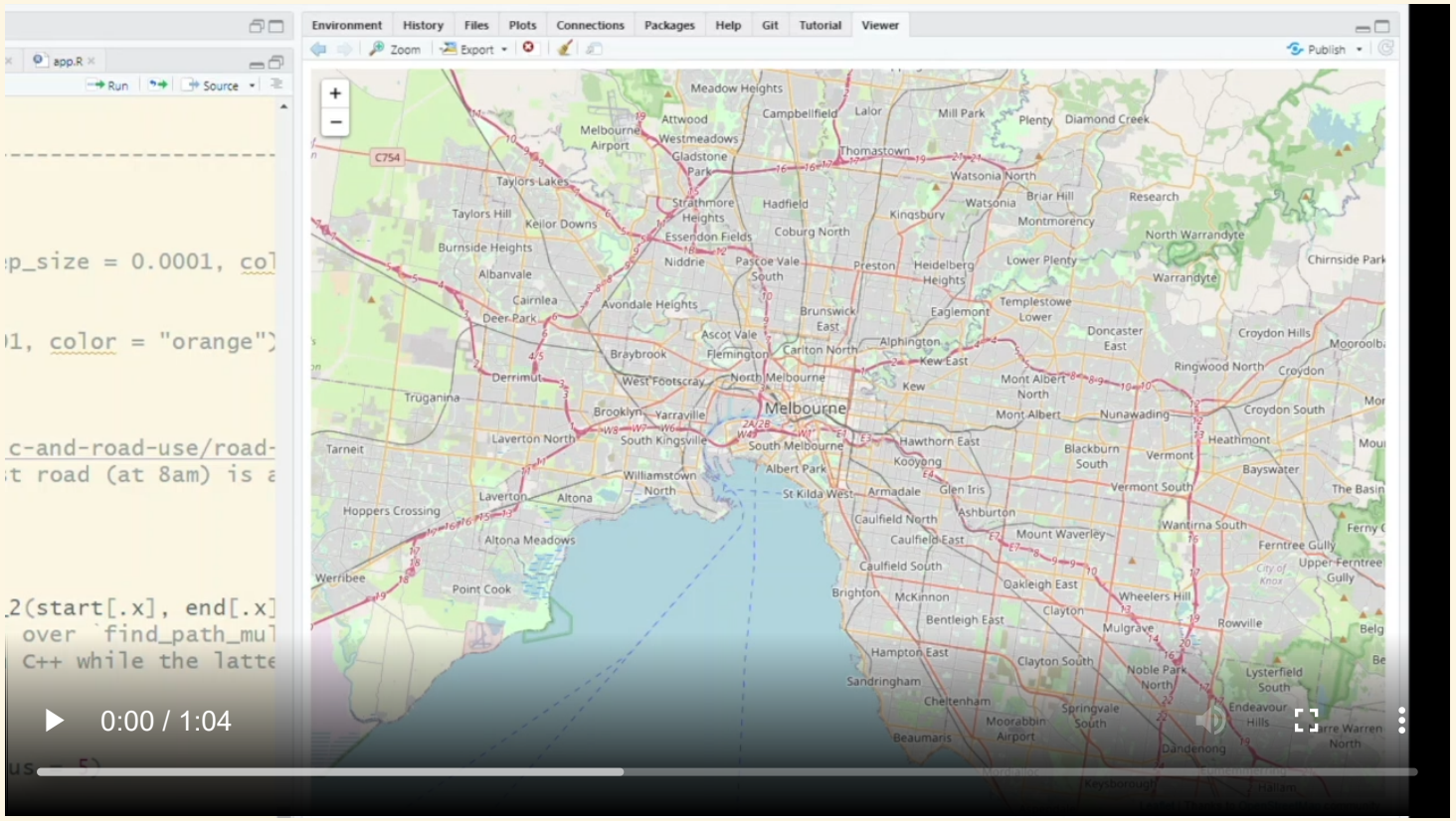
Enhancing the model with web mapping

Suppose an agent decides to move from suburb s to suburb s',

We can sample one street node from each of s and s', and make the agent travel the path connecting the two nodes.

- Macro-level numbers remain exactly the same.
- Extra freedom to model the micro-level movement using another model (statistical or mechanistic).

Demo


Environment
History
Files
Plots
Connections
Packages
Help
Git
Tutorial
Viewer
Zoom
Export
Publish
app.R
Run
Source
p_size = 0.0001, col
1, color = "orange")
c-and-road-use/road-
t road (at 8am) is a
2(start[.x], end[.x]
over `find_path_mul
C++ while the latte
0:00 / 1:04
Melbourne
Meadow Heights
Attwood
Campbellfield
Lalor
Mill Park
Plenty
Diamond Creek
Melbourne Airport
Westmeadows
Gladstone Park
Thomastown
Watsonia North
Taylors Lakes
Strathmore Heights
Hadfield
Watsonia
Briar Hill
Research
Taylors Hill
Keilor Downs
Kingsbury
Montmorency
Coburg North
Burnside Heights
Essendon Fields
North Warrandyte
Albanvale
Niddrie
Pascoe Vale South
Preston
Heidelberg Heights
Lower Plenty
Chirnside Park
Warrandyte
Cairnlea
Deer Park
Avondale Heights
Brunswick East
Eaglemont
Templestowe Lower
Ascot Vale
Flemington
Carlton North
Alphington
Doncaster East
Croydon Hills
Braybrook
Kew East
Ringwood North
Croydon
Derrimut
West Footscray
North Melbourne
Kew
Mont Albert North
Truganina
Brooklyn
Yarraville
Mont Albert
Nunawading
Croydon South
Laverton North
South Kingsville
South Melbourne
Hawthorn East
Blackburn South
Heathmont
Tarneit
Albert Park
Kooyong
Vermont
Bayswater
Williamstown North
Vermont South
Laverton
Altona
St Kilda West
Armadale
Glen Iris
Ashburton
Hoppers Crossing
Caulfield North
Wantirna South
Altona Meadows
Caulfield East
Mount Waverley
Ferntree Gully
Caulfield South
Oakleigh East
Werribee
Point Cook
Brighton
McKinnon
Wheelers Hill
Bentleigh East
Clayton
Mulgrave
Rowville
Hampton East
Clayton South
Noble Park North
Lysterfield South
Sandringham
Cheltenham
Springvale South
Endeavour Hills
Moorabbin Airport
Beaumaris
Dandenong

Summary

Key idea

Combine web map data, data analysis / simulation in R and interactive web map display to answer spatial and map queries.

Use cases

- ✔ Make basic queries and interactively manipulate the map
- ✔ Improve the spatial resolution of a Markov chain on graphs
- 💡 Planning / evaluating road block disrruption
- 💡 Optimal placement of pedestrian counters
- 💡 Creating location profiles

References

1. OSM data available at Geofabrik https://download.geofabrik.de/ or BBBike: https://download.bbbike.org/osm/bbbike/Melbourne/

2. Australian map tagging guidelines: https://wiki.openstreetmap.org/wiki/Australian_Tagging_Guidelines

3. The `sketch` R package https://github.com/kcf-jackson/ecssc-2021-talk/ and the `leaflet` JavaScript library https://leafletjs.com/.

I work at the Bioinformatics and Cellular Genomics lab (BioCellGen) @ St. Vincent's Institute of Medical Research (SVI) with interest in statistical computing and visualisation.

Talk repo at: https://github.com/kcf-jackson/ecssc-2021-talk/

Feel free to reach me at jkwok@svi.edu.au or via Slack on ROpenSci.