Towards a city congestion index: methodological explorations using Google's Distance Matrix API

Jonathan Cohen and Jorge Gil

Chalmers University of Technology, Sven Hultins gata 6, SE-412 96, Göteborg, Sweden

E-mail: `jonathan.cohen@chalmers.se`

Abstract. All articles *must* contain an abstract. This document describes the preparation of a conference paper to be published in *Journal of Physics: Conference Series* using LaTeX 2_ε and the `jpconf.cls` class file. The abstract text should be formatted using 10 point font and indented 25 mm from the left margin. Leave 10 mm space after the abstract before you begin the main text of your article. The text of your article should start on the same page as the abstract. The abstract follows the addresses and should give readers concise information about the content of the article and indicate the main results obtained and conclusions drawn. As the abstract is not part of the text it should be complete in itself; no table numbers, figure numbers, references or displayed mathematical expressions should be included. It should be suitable for direct inclusion in abstracting services and should not normally exceed 200 words. The abstract should generally be restricted to a single paragraph. Since contemporary information-retrieval systems rely heavily on the content of titles and abstracts to identify relevant articles in literature searches, great care should be taken in constructing both.

1. Introduction

What are the problems with congestion? Is it mentioned in UN Habitat NUA, or SDGs e.g.11?

What are the clients of this index? Who needs such an index, and to do what?

2. Background

What congestion indices exist out there?

Practice/literature

How are others mapping congestion (or real travel time) with geodata? (edited)

3. Research Aim

As mentioned above, traffic congestion is an urban problem that deteriorate the positive benefits of urban life, therefore is imperative to understand the problem in depth and look at it from different angles. Traditional congestion measurements are found to be costly and through practice or research different methods have been explored to overcome this issue.

One avenue of research understands congestion as a time and location specific problem, which occurs in a road segment and that its consequences are suffered from those in the immediate surroundings. A second understanding of the problem deals with the individuals who are affected by traffic.

Focusing on this second interpretation of the problem, the aim of this research is to provide a methodology to spatialize congestion to expose how traffic congestion is distributed across an urban area. The map of congestion can provide useful information to local authorities and planners as

spatial information about the city and consequently provide and measure how the problem has been evolving over time and space.

4. Methodology

At its core the methodology described in this section details a process of collecting and processing data from an internet service that provides traffic estimations and routing optimization. The method presented here correspond to data extracted from the Distance Matrix API form Google Inc., but the exact same exercise could be done with other services such as HERE or TOMTOM.

Given the quantity and nature the methodology that will be described above, was coded and executed using R in Rstudio and the functions written can be found in GitHub.

In order to prioritize clarity over narrative, the necessary steps followed by this methodology are listed below:

1) Select and define an urban area to work with: After an urban area is selected a synthetic boundary needs to be drawn. In this case, for simplicity reasons, a circular buffer zone was used to define the 'study area'.
2) Choose a zoned cartography of the area and clip it accordingly to the boundaries: Choose from any geographical authority a zoned cartography. Census tracks, neighbourhoods or plots can be chosen. In our case the European Population grids was selected.
3) Make centroids from the zones and extract the geographical position of each point
4) Generate permutations from centroids to create an empty Origin-Destiny Matrix: Using the GPS location (latitude and longitude) of each centroid, a list of all possible combinations is generated (a total of n*(n-1) routes). For instance, is the area to be studied contains 3 areas (A, B & C), the OD matrix will consider the following routes: (i) A to B, (ii) A to C, (iii) B to A, (iv) B to C, (v) C to A & (vi) C to B.
5) Define a *congested* and *non-congested* times: Using the difference between a *congested* and *non-congested* situation as the definition of congestion, a time need to be specified. As information from Google will be used, a future time needs to be specified. Based on previous results, Thursday, October 15, 2020 8:30:00 AM was chosen as a *congested* moment and Wednesday, July 15, 2020 3:30:00 AM, as a *non-congested.*
6) Use a routing service to estimate time and distance travelled: The origin to destination list with the *congested* and *non-congested* time specification were used as inputs for the routing service to estimate travel time and distance.
7) Calculate descriptive statistics and visualize: Histograms, scatter plots and descriptive statistics were calculated to detect possible problems such as missing data or outliers.
8) Deal with data issues: As the centroid of the polygon based on the European Population Grid could fall over a problematic place such as river, park or railway lot (places where google maps cannot match an address), some routes could not be retrieved and consequently were left out of the study.
9) Calculate KPIs: Based on research done in the past the following can be derived:
   - Time difference:
   - Relative time difference:
   - Average travel speed:
   - Average travel speed difference:
   - Relative travel speed difference:

- Travel Time Index (TTI):

10) Aggregate data generated and join to spatial grid: With the data retrieved averages by the origins were calculated and joined to these values joined to the original zoned urban area.
11) Calculate descriptive statistics and visualize: The results at the Grid level are analyzed using descriptive statitics and basic plotting. This step can be seen as check-up point that can contribute to identify data anomalies such as missing data or outliers.
12) Map the results: Finally, once the data clean-up process is done, a map with the processed variables can be generated.

To demostrate how the method could be used in practice, four European cities where selected as a proof of concept.

5. Data

In order to run the analysis for a desired urban area, a zoned cartography of the area is needed. In this opportunity, as a proof of concept the 1km2 population grid from the EuroStat was used. Amsterdam, Glasgow, Goteborg and Lisbon were selected and from their historical centre a buffer zone was used to delimit the *city* boundary. The geographic coordinates of the centroid of each 1km2 square was used to create the list of all possible routes.

The list of all possible origins to all possible destinations was then used to retrieve data for a *congested* and *non-congested* moments.

5.1. Data retrieved

The process generates information for a total n*(n-1) routes (n being the number of zones). For instance, Lisbon is divided into 119 squares and the number of possible routes is 119*118 = 14,042 (then 28,084 travels were estimated). Amsterdam has 131 zones, Glasgow 136 and Goteborg 123 (17,030, 18,360 and 15,006 routes respectively). A summary of the descriptive statistics for the time in minutes can be found in Table 1.

	Congested scenario (mins)	Routes	Mean	S.D.	Min	Max
1	Amsterdam	17,023	20.26	6.86	0.65	48.57
2	Glasgow	18,355	23.08	8.68	1.42	56.88
3	Goteborg	14,997	18.01	6.82	1.37	57.97
4	Lisbon	14,043	28.16	11.70	1.37	84.63
	Non-Congested scenario	Routes	Mean	S.D.	Min	Max
1	Amsterdam	17,023	11.70	3.71	0.52	24.90
2	Glasgow	18,355	11.93	3.99	0.83	24.85
3	Goteborg	14,997	13.09	5.29	1.10	49.18
4	Lisbon	14,043	12.82	5.77	0.77	46.90

Table 1: Descriptive Statistics of travel times by city

The amount of route information retrieved is consistent with the amount of zones in the city and as expected, the *congested* situation on average takes longer time. Formally an independent-samples t-test was conducted to compare the travel times under a *congested* and *non-congested* situations. For each city, the time difference was found statistically significant with a confidence level of 99%.

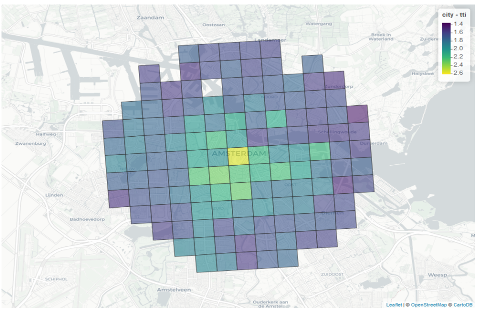
6. Results

After capturing, processing and aggregating the data obtained from the API, table 2 on page 4 contains the descriptive statistics for time difference and TTI. In this case, the number of observations corresponds to the number of zones in the map and each zone holds the mean of all the trips departed from that origin. The results show that the methodology successfully captured diversity between cities. For instance, in Lisbon the average time difference is of 15.3 mins, with a maximum of 21.6 mins, meanwhile in Goteborg on average delays are of 4.9 mins with a maximum of 8 mins.

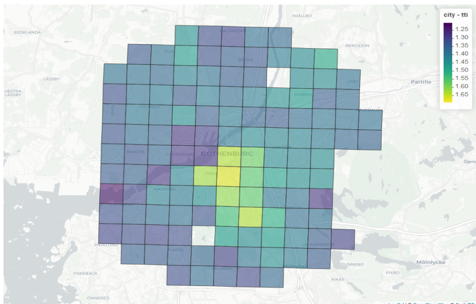
	Time Difference (mins)	Zones	Mean	S.D.	Min	Max
1	Amsterdam	131	8.57	2.39	5.38	19.03
2	Glasgow	136	11.15	1.88	7.66	15.64
3	Goteborg	123	4.92	0.99	3.22	8.06
4	Lisbon	119	15.34	2.67	9.74	21.57
	TTI	Zones	Mean	S.D.	Min	Max
1	Amsterdam	131	1.74	0.23	1.36	2.61
2	Glasgow	136	1.94	0.14	1.60	2.31
3	Goteborg	123	1.38	0.09	1.21	1.69
4	Lisbon	119	2.21	0.16	1.70	2.57

Table 2: Aggregated descriptive statistics by zones

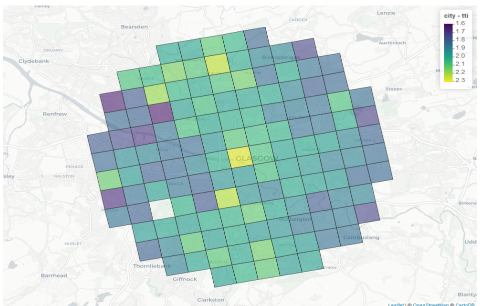
Although, the table above reveals relevant insights about these cities, local authorities or planners would not know which places or who is being affected by traffic congestion. Therefore, in figure 2 shows a map for each city with the values of TTI for each zone.



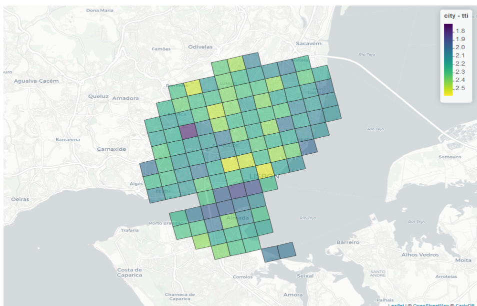
(a) Amsterdam



(c) Goteborg



(b) Glasgow



(d) Lisbon

Figure 1: Images extracted from results

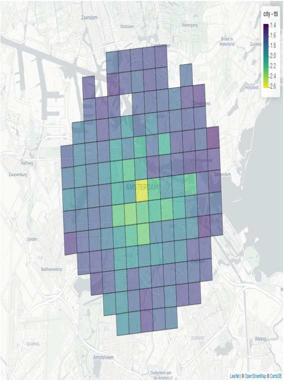
7. Discussion

* The methodology presented in this paper

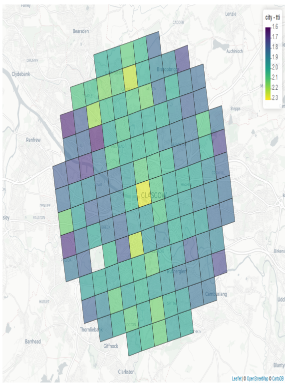
7.1. Scalability

7.2. Limitations and future work

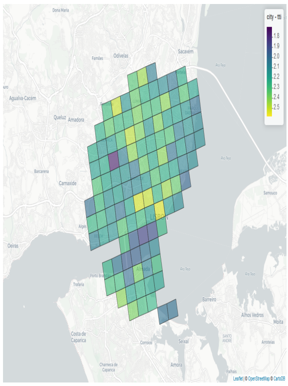
8. Conclusion



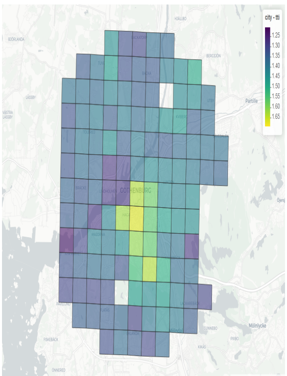
(a) Data Science



(b) New Data Sources



(c) Data Capture



(d) Visualization

./Img/2_1.PNG

(e) Data Capture

./Img/2_2.PNG

(f) Visualization

./Img/2_3.PNG

(g) Data Capture

./Img/2_4.PNG

(h) Visualization

./Img/3_1.PNG

(i) Visualization

./Img/3_2.PNG

(j) Data Capture

./Img/3_3.PNG

(k) Data Capture

./Img/3_4.PNG

(l) Visualization

./Img/4_1.PNG

(m) Visualization

./Img/4_2.PNG

(n) Data Capture

./Img/4_3.PNG

(o) Visualization

./Img/4_4.PNG

(p) Visualization