Tracing tourism geographies with Google Trends: A Dutch case study

Andrea Ballatore, Simon Scheider, and Bas Spierings

Author copy, 2019
In: Kyriakidis, P., Hadjimitsis, D., Skarlatos, D. & Mansourian, A. (eds.), AGILE Conference on Geographical Information Science: Geospatial technologies for local and regional development. Berlin: Springer.

Abstract Search engines make information about places available to billions of users, who explore geographic information for a variety of purposes. The aggregated, large-scale search behavioural statistics provided by Google Trends can provide new knowledge about the spatial and temporal variation in interest in places. Such search data can provide useful knowledge for tourism management, especially in relation to the current crisis of tourist (over)crowding, capturing intense spatial concentrations of interest. Taking the Amsterdam metropolitan area as a case study and Google Trends as a data source, this article studies the spatial and temporal variation in interest in places at multiple scales, from 2007 to 2017. First, we analyze the global interest in the Netherlands and Amsterdam, comparing it with hotel visit data. Second, we compare interest in municipalities, and observe changes within the same municipalities. This interdisciplinary study shows how search data can trace new geographies between the interest origin (what place users search from) and the interest destination (what place users search for), with potential applications to tourism management and cognate disciplines.

Key words: Interest geography; Place search; Web science; Google Trends; Tourism; Amsterdam; Netherlands

Andrea Ballatore (✉)
Department of Geography, Birkbeck, University of London, London, UK
e-mail: a.ballatore@bbk.ac.uk

Simon Scheider
Human Geography and Planning, Universiteit Utrecht, Utrecht, NL
e-mail: s.scheider@uu.nl

Bas Spierings
Human Geography and Planning, Universiteit Utrecht, Utrecht, NL
e-mail: b.spierings@uu.nl

1 Introduction

In their effort to provide relevant information to users, search engines create connections between users and Web content. Just considering Google, more than a billion people query the search engine every day, resulting in more than 3.5 billion requests per day.[1] Geography plays a central role in this process of information retrieval and consumption, as users, machines, and online resources are always located somewhere on the planet (Graham et al., 2015). Users search for places from their location for a variety of purposes, including travel, reference, news, entertainment, and investment. The expression of interest in an entity (e.g. typing "London" into Google) is a relatively novel and unique type of big data that can be used for the study of human dynamics (Stephens-Davidowitz, 2013).

Tourism constitutes a major and increasing driver for visiting places. Over the past decade, global tourist flows have strongly increased, exhibiting a growth rate of 4% or more every year. In 2017, the growth was even 7% reaching a total of 1.32 billion overnight visitors worldwide.[2] Search engines, combined with online platforms such as booking.com and Airbnb, have radically transformed how tourists find their destinations and plan their journeys. The economic and cultural benefits of growing tourism do not come without a cost. Many focal areas in the world suffer from overcrowding, including Venice, Barcelona, and Amsterdam. In this context, big data analytics offer new insights into observing, modelling, and forecasting tourism-related behaviour in space and time (Önder and Gunter, 2016).

As part of the broad "Web science" approach (Hendler et al., 2008), the potential of search data has attracted attention from multiple viewpoints. Since 2006, Google Trends[3] has provided aggregated measures of interest in topics searched for on the search engine, capturing how interest in a person, city, event, book, team, or movie can rise, fall, concentrate, and disperse over time (Jun et al., 2018). This data source spurred several applications in economics (Choi and Varian, 2012) and epidemiology (Lazer et al., 2014). While some of these works include geographic and temporal dimensions, no study has focused on the search for places with an explicit attention to the spatial structure of this human behaviour.

In this article, we use Google Trends data to observe the variation of interest in places, with particular attention to potential applications in tourism studies. This exploratory study of what we might call an "interest geography" aims at paving the way towards more modelling-oriented efforts. As a case study, we consider the Amsterdam metropolitan area, a region of the Netherlands that comprises 33 municipalities. The investigation is restricted to the most recent decade with complete data (2007–2017). Central to the study are the interest origin (which place people search from) and the interest destination (which place people search for). Following

[1] https://web.archive.org/web/20181204153621/http://www.internetlivestats.com/google-search-statistics. All URLs were accessed in November 2018, and are stored in the Internet Archive.

[2] https://web.archive.org/web/20180808124718/http://cf.cdn.unwto.org/sites/all/files/pdf/unwto_barom18_01_january_excerpt_hr.pdf

[3] https://web.archive.org/web/20181206023025/https://trends.google.com/trends/

recommendations by Singleton et al. (2016), all analyses in this article are fully reproducible, and code and data are available online.[4] This article is a first step towards answering the following research questions:

RQ1 How does search behaviour change at different geographic scales (e.g. national, regional, local)?
RQ2 How does search behaviour change in space across different interest origins and different interest destinations?
RQ3 Is there any correlation between search behaviour and measurable tourism activity, such as hotel visits?
RQ4 What areas are over- and under-represented in online search interest with respect to population size?

The remainder of this article is organized as follows. Section 2 discusses related work in spatio-temporal analytics of search data, particularly on Google Trends, and tourism studies. The analysis of interest origin for the Netherlands and Amsterdam at the global scale is then discussed in Section 3, contrasting hotel visits with search interest. Section 4 analyzes the variation of interest in municipalities in the Amsterdam metropolitan area over time, both within each municipality and between municipalities. Section 5 concludes the article with a discussion and outlook on the potential of search data for tourism analytics.

2 Related work: Search interest and tourism

This study is located at the intersection of web science, internet geography, data science, and tourism studies. This section aims at covering relevant works from these areas, including background about tourism in the Amsterdam metropolitan area.

2.1 Search engine data for human dynamics

The information retrieval community has traditionally relied on search engine logs as a source of information about user behaviour, before privacy concerns emerged during the notorious AOL incident in 2006.[5] Since then, search data has been released only in coarse aggregated forms, making it impossible to connect searches with specific individuals. Additionally, as the business model of Google relies on the secrecy of its algorithm to reduce ranking manipulation in the arena of Search Engine Optimization (SEO), access to highly granular search data is essentially impossible outside of the search engine provider.

[4] https://github.com/andrea-ballatore/SearchGeography
[5] https://web.archive.org/web/20181204192954/https://www.nytimes.com/2006/08/09/technology/09aol.html

For these reasons, Google Trends indicates an index of interest from 0 to 100, calculated relative to *all searches* conducted in a given period, allowing comparisons either between terms, or within the same term over time, and does not provide the actual number of searches. Therefore, if the overall pool of searches grows and changes composition, the index for a given topic can *decrease* even if the actual number of searches has increased. Given the noisy, volatile, and ambiguous nature of search data, caution is needed when handling the Google Trends index.[6]

Search motivations vary widely (e.g. searching for London to plan a journey or to learn about its history), and the disaggregation of these behaviours is not trivial. Search terms are often ambiguous ("Chelsea" as an English or American neighbourhood, an English football team, or a female name), and exceptional events, such as natural and man-made catastrophes, tend to generate short-lived bursts of interest. Spatially, the geography of searches is biased by an uneven user distribution: Google has a limited user base in major countries such as China, Russia, and Iran. Furthermore, virtual private networks (VPNs) deliberately obfuscate the geographic location of users.

Despite these limitations and challenges, Google Trends data have been used in a variety of contexts, primarily in economics (Jun et al., 2018) and management (Askitas and Zimmermann, 2015). Researchers showed that search data can forecast future behaviours (Goel et al., 2010), often correlate with changes in the value of several goods and services (Choi and Varian, 2012), and can quantify attitudes that are not easily revealed by polls, such as racism (Stephens-Davidowitz, 2013). From a more geographical perspective, Google Flu Trends famously linked search locations to potential flu outbreaks, producing some hype and subsequent reappraisal of the potential of big search data for forecasting (Lazer et al., 2014).

In GIScience and geography, our prior work focuses on several facets of search engines, including the localness of results (Ballatore et al., 2017), interest in crowdsourced datasets (Ballatore and Jokar Arsanjani, 2018), and popularity of data science tools (Ballatore et al., 2018). However, none of these studies has the spatial focus that we take in this article.

2.2 Tourist (over)crowding and big data

Cities and their historical cores have since long been popular recreational destinations among travellers (Gospodini, 2006), but only since the end of the 20th century, assets are purposefully developed and marketed in order to create and foster a "visitor economy" (Hall and Barrett, 2017). Ambitious city marketing strategies were often developed and opportunistic city development projects implemented to sell and boost the city's appeal for visitors in terms of leisure and consumption, with the goal of attracting large numbers of visitors and their spending power (Spierings, 2013). Supported by, amongst others, the rise in leisure time, the growth of spending

[6] https://web.archive.org/web/20181108133214/https://medium.com/@pewresearch/using-google-trends-data-for-research-here-are-6-questions-to-ask-a7097f5fb526

power, and the proliferation of low-cost airlines, many cities have also become very successful over time in attracting tourists. However, rising complaints and protests by residents against large and increasing numbers of tourists are clear indications that several cities have recently become *too* successful, and are increasingly perceived by residents and tourists alike as "(over)crowded" (Popp, 2012; Novy and Colomb, 2017; Neuts and Vanneste, 2018).

Even though surveys provide valuable quantitative data about tourist flows and related processes of (over)crowding, supporting decision-making, the rise of "big data" promises cheaper and more granular data sources (Kitchin, 2014). The potential of search engine behaviour has not gone unnoticed in tourism and management studies (Pan et al., 2012; Yang et al., 2015). Search statistics can provide valuable information about the interests and intentions of tourists (Li et al., 2017). The estimation of tourist arrivals can also be enriched by search trends, as shown in a case study on Vienna (Önder and Gunter, 2016). Search interest correlates better with arrivals when linguistic and other biases are taken into account (Dergiades et al., 2018). Along similar lines, Siliverstovs and Wochner (2018) improved the forecasting accuracy of Google trends for tourist arrivals to Switzerland by employing a cross-sectional instead of a longitudinal approach. This body of work offers sophisticated insights into the temporal dimension of search patterns, but overlooks the spatial dimension that we focus on in this article.

2.3 Tourist marketing and spreading in Amsterdam

As our case study focuses on the Amsterdam metropolitan area, it is beneficial to provide background about the city's complex relationship with tourism. Nowadays, most tourist destinations aim to manage and mitigate the social impacts of (over)crowding. One of the strategies that is experimented with is the spatial "spreading" of tourists. Amsterdam is a telling case of this policy strategy, which has the objective of spreading tourists at the urban, regional, and even at the national level. Recently, a marketing campaign was initiated to promote neighbourhoods outside and adjacent to the historical core of the city as interesting and diverse destinations.[7]

A complementary tourist-spreading strategy was developed at the national level with Amsterdam as an important constituent of several marketing "storylines". These narratives present the country as one metropolis (i.e. Holland city) and promote visits "off the beaten track" throughout the country.[8] The storyline of the Dutch Golden Age, for instance, combines the well-known capital's canals and the Rijksmuseum with attractions in cities like Haarlem, Leiden, Middelburg, and Dordrecht, inviting tourists to go beyond Amsterdam.

---

[7] https://web.archive.org/web/20181206203125/https://www.iamsterdam.com/nl/over-ons/amsterdam-marketing/afdelingen/marketing-strategy/consumer/buurtencampagne

[8] https://web.archive.org/web/20181206203356/https://www.nbtc.nl/en/homepage/collaboration/storylines.htm

A similar marketing campaign has been running for more than a decade.[9] Several places and sites within the urban region are marketed as extensions of Amsterdam – including Zandvoort aan Zee (a seaside resort) as part of "Amsterdam beach", Muiderslot (a medieval castle in Muiden) as "Amsterdam castle", Zaanse Schans and Volendam (respectively a traditional village with windmills and a traditional fishing village) as "old Holland", and Almere (a city developed in the second half of the twentieth century) as "new land". Our study of search geography brings a new viewpoint on how to sense and gather data about the spatial configuration of tourist interest.

3 Interest origin for the Netherlands and Amsterdam

To inspect the geography of search, we start by observing the interest origin (i.e. where people search from) for the Netherlands and Amsterdam, two prominent interest targets in our Dutch case study. Hence, we collected the Google Trends index (GTI) from 2007 to 2017 at the country level.[10] For example, the GTI for Amsterdam in France in 2010 is 2, with 100 being the GTI in the country that showed the highest interest (the Netherlands). In order to gain an understanding of how many searches actually occurred, we obtained estimates from SEO company SemRush. The company indicates that, in October 2018, "Amsterdam" received 246,000 searches from the Netherlands and 201,000 from the UK. These estimates are useful to sense the magnitude of the search volume.

During a first inspection of the GTI, we noted an unusual peak in July 2014 for "Netherlands", corresponding to the World Cup. It is therefore highly likely that most of those searches were aimed at the Dutch national football team and not at the country. "Amsterdam", by contrast, showed seasonal variation without bursts. Hence, in the analyses below we focused on Amsterdam as opposed to the Netherlands, for which correction would be needed. To reduce the noise in the data, unless specified otherwise, we retrieved Google Trends data with the "low volume" option off. This option allows to include or exclude data points with low values that are filtered out by default.

In Google Trends, it is possible to retrieve data as a "term" (a plain string), or as a "concept" (a bundle of correlated terms). While, in principle, the results can vary significantly between terms and concepts, empirical observation indicated that the variation was negligible in this study, and therefore we always used terms. As the difference in interest for objects of different sizes can vary by orders of magnitude, when appropriate, we show transformed values with inverse hyperbolic sine, conceptually similar to a logarithmic transformation, but better at handling 0 values.

[9] https://web.archive.org/web/20181206102019/https://www.iamsterdam.com/en/plan-your-trip/day-trips

[10] The data was retrieved in tabular form from the Google Trends API using the R package *gtrendsR*: https://cran.r-project.org/web/packages/gtrendsR

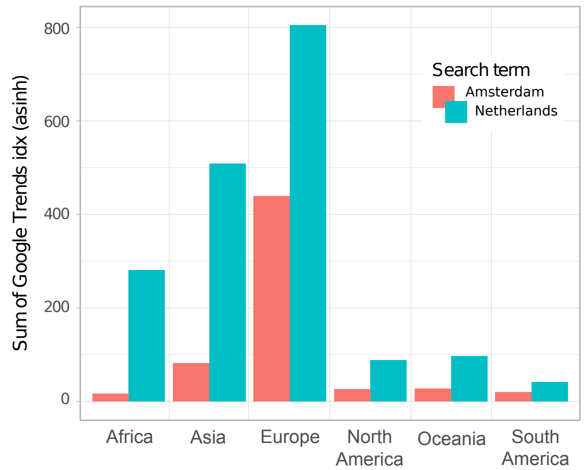

Fig. 1: Google Trends interest (GTI) for terms “Amsterdam” and “Netherlands”, summing countries by continent from 2007 to 2017 (GTI transformed with inverse hyperbolic sine, without low-volume data).

3.1 Spatial distribution of interest origin

The GTI quantifies the interest for the interest destinations (Amsterdam and the Netherlands) in each country in the time period, relative to other countries, which makes the comparison between countries possible. As a first step to describe the overall geography of interest origin, we aggregated the countries by continent (omitting Antarctica), summing the GTI for the 11 years included in the dataset, as shown in Figure 1. For all continents, the Netherlands obtained higher interest than Amsterdam, suggesting a proportionality between the size of the interest target and the GTI (RQ1). Europe shows the highest GTI for both terms by far, with the Netherlands obtaining 1.7 times the summed GTI than Amsterdam. Notably, Africa obtained 17.6 times more interest in the country than in the capital, while the same ratio for Asia is 6.2. In this sense, searches from Africa focus disproportionately on the whole country as opposed to the capital city. The other continents show substantially lower values, indicating a strong distance decay effect. At a closer look at European countries, the Netherlands generates the highest GTI (100) for both terms, followed by Belgium, UK, Ireland, Germany, and Italy.

At a finer spatial granularity, Figure 2 shows the variation in GTI by country for Amsterdam between 2007 and 2017 (RQ1). Note that for this comparison, low-volume data was included, as most African countries would have been missing. These maps clearly show that interest origin became less diverse and more concentrated in Europe over time (RQ2). A possible explanation for this change is that, in the decade being analyzed, European tourist flows grew comparatively more than those from other continents (see Section 3.2). Beyond the striking dominance of Europe,

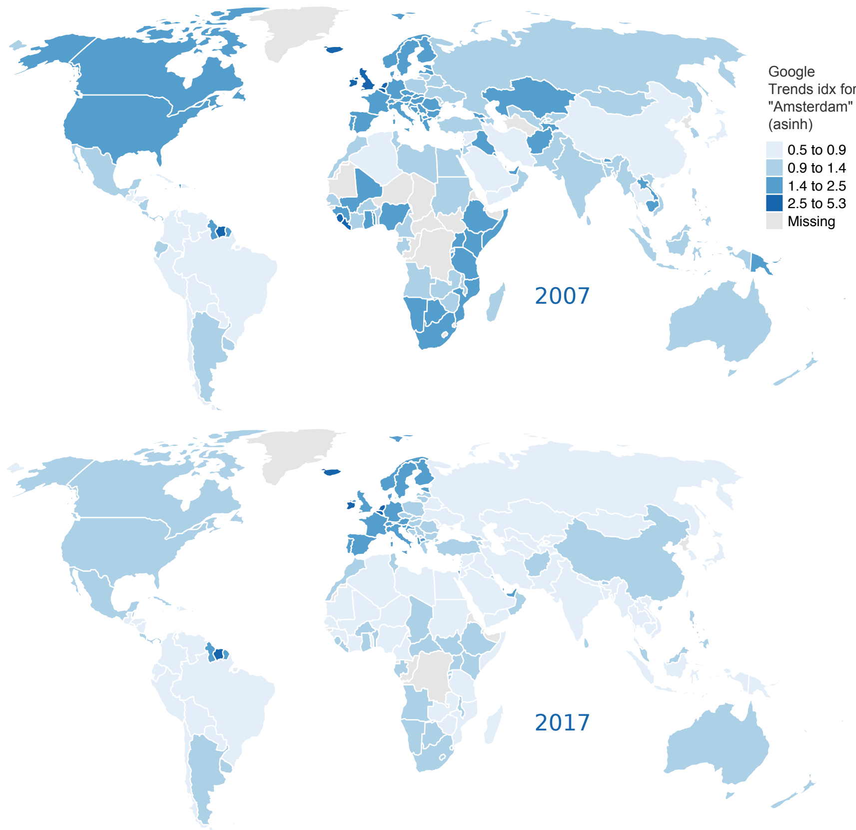

Fig. 2: Google Trends interest for terms “Amsterdam” by country 2007 and 2017, with low-volume data (index transformed with inverse hyperbolic sine). Data grouped with quantiles. Period 2007–2017. Projection: Robinson. World borders from Natural Earth, 2017.

post-colonial links emerged. Countries with unusually high interest (GTI in range [8,23]) include former or current Dutch territories, such as Suriname, Aruba, Curaçao, and Sint Maarten. By contrast, Liberia also has an unusually high GTI, whereas colonial relations with the Netherlands were very short-lived.

3.2 Temporal change in interest and hotel visits

To study the relationship between quantifiable tourist behaviour and search data (RQ3), we retrieved hotel visits in Amsterdam from 2012 to 2017 for 50 countries.[11]

[11] https://web.archive.org/web/20181201124839/http://statline.cbs.nl/Statweb

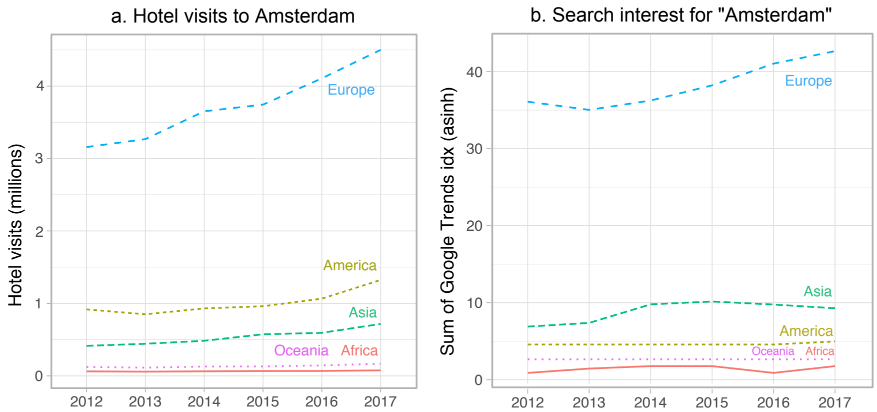

Fig. 3: (a) Hotel visits in Amsterdam and (b) GTI for term "Amsterdam" from 2012 to 2017, without low-volume data (index transformed with inverse hyperbolic sine) for 50 countries. Sources: Statistics Netherlands and Google Trends.

This data estimates stays in traditional hotels, and does not include home sharing platforms, such as Airbnb. Figure 3 depicts the yearly change in hotel visits and the search interest as the sum of transformed GTI per continent in the same period, for the 50 countries included in the dataset. It is possible to note that a superficial similarity between the trend lines, particularly for the case of Europe, dominant and increasing both in terms of hotel visits and GTI. However, the trends diverge for other continents, which show a slight decline in GTI, while hotel visits actually increased, even though not at the same pace as Europe (RQ3).

The number of years in the dataset (6) is too low for meaningful correlation analysis at the country level, we calculated the pair-wise correlation of all countries for all years. The hotel visits are highly granular (181 levels in the variable), but because of the coarse granularity of GTI (the variable takes only 12 levels in the case of Amsterdam), the pairs contain a large number of ties, and therefore the non-parametric Kendall's τ is a suitable correlation coefficient. Table 1 shows the global pair-wise correlation between hotel visits and GTI. No meaningful correlation is found for the Netherlands (τ near 0), confirming the difficulty of extracting tourism-related value from Google Trends with a query affected by interest bursts (e.g. the World Cup). By contrast, a strong positive correlation is found for Amsterdam, without low-volume data ($\tau = .57$).

Over time, this correlation appears to have decreased (from .63 in 2012 to .51 in 2017, at $p < .001$), and we have no clear interpretation of this fact. Low-volume data shows weak correlations in both cases. This indicates that low-volume data is noisier and needs extra caution to be handled. Bearing in mind the specificity of the case study, these results suggest that searches for cities correlate fairly well with hotel visits, while no correlation is visible with searches for countries (RQ3). Intuitively,

Target	Low-volume	GTI levels	Kendall's τ
Amsterdam	No	12	.57***
–	Yes	11	.13**
Netherlands	No	32	-.03
–	Yes	25	-.13**

Table 1: Correlations between hotel visits and GTI in Amsterdam and in the Netherlands (2012–2017). $N = 294$ for each test. Significance level: (***) $p < .001$, (**) $p < .01$, (*) $p < .05$.

this might be explained with the fact that hotel are searched for in specific cities, and not in whole countries.

4 Spatial analysis of interest in municipalities

After studying the global interest origin for Amsterdam and the Netherlands, we turned our attention to municipalities (*gemeenten* in Dutch) as interest destinations in the Amsterdam metropolitan area, which covers about 2,600 km^2 and currently hosts 2.3M residents. Its 33 spatial administrative units represent Amsterdam and the cities surrounding it. The largest municipalities in terms of resident population are Amsterdam (about 845,000), Almere (201,000), and Haarlem (159,000), while the smallest municipalities have as few as 6,000 residents.

4.1 Data granularity and comparability

Retrieving data about these municipalities from Google Trends involves distinct technical challenges. The GTI is computed relative to a set of up to 5 search targets. Since the index is expressed as a percentage of the maximum amount of searches over all targets, indices from different queries are not comparable (for example, GTI = 10 for Amsterdam would indicate a lot more searches than the same GTI for a less popular place). To overcome this limitation, we included a reference term in each query which was guaranteed to define the maximum for all target objects, namely "Amsterdam". This effectively made all municipality indices comparable with the reference term, and thus comparable to each other. A second strategy was to compare the temporal behaviour between regions, in terms of decline or increase in percentage.

A second difficulty lies in the fact that the GTI has a very limited resolution. GTI values near zero are rounded to zero, and all other values are rounded to integers. This limitation is enforced by design to avoid exposing fine-grained information that could be used to alter rankings. This makes it difficult to compare index distributions over interest targets that are spread out with a long tail in higher values, because many low values are reduced to zero or 1.

In order to collect data at a sufficient granularity, we implemented an algorithm that increases the resolution of a given query. This is done by subdividing the query terms along their order of magnitude at GTI gap of more than 80, then re-querying the lower subset of terms, and finally scaling the new index values of the subset to the old index value. For example, one can query the GTI for “Amsterdam”, “Almere” and “Beemster” for 2018. The result is Amsterdam = 100 (base term), Almere = 6, and Beemster < 1. Since the GTI gap is more than 80, a second query is executed with Almere as the base term (100), and Beemster obtains 2. We can then estimate the GTI for Beemster as $2/100 \cdot 6 = 0.12$. While this procedure can in principle be repeated for all index gaps, we only applied it once, since the differences between terms of lower interest turned out to be smaller than 80.

For each municipality, we inspected Google Trends and ensured that the query was fairly unambiguous, i.e. the municipality name did not have another obvious meaning. Interestingly, smaller municipalities were under-represented in Google Trends, with undefined GTI in most instances (RQ1). This suggests that usable data exists for countries, cities, and for some popular neighbourhoods and points of interest (e.g. Van Gogh Museum), but smaller areas are not mappable.

Using the aforementioned algorithm, we collected the GTI *between* municipalities, for period 2007–2017, using a base term to make them comparable at a given time (e.g. was the interest higher for Amsterdam or for Almere in 2007?). The data was then collected *within* municipalities, looking at the temporal variation for a single municipality (e.g. was the interest for Almere higher in 2007 or in 2008?). In both datasets, the GTI excluded low-volume data.

4.2 Interest between municipalities

In this analysis, we used the aforementioned algorithm to be able to compare the interest in target municipalities across space, starting from “Amsterdam” as a base term (GTI $= 100$). Note that all the following maps of the Amsterdam metropolitan region are projected with UTM (zone 31N). The boundaries and population data are from dataset *CBS Wijk- en Buurtkaart 2017*.[12] Figure 4 shows the GTI of the 33 municipalities, ranging from 0 to 62.5.

As can be seen, high-interest muncipalities emerged in the suburban area (RQ2). On the Western coast, Zandvoort is a popular beach area close to Amsterdam. Nearby, Haarlem is an important historical city with several popular museums. Amstelveen has an important residential function for people working in Amsterdam, but also contains corporate head offices, including that of KLM, and a prominent art museum. Huizen is a former fishing village. Hilversum is known as the Dutch “media city” with a focus on radio and television broadcasting and a related museum on image and sound. Finally, Almere and Lelystad were both developed in the second half of the 20th century, and are currently important centers for commuters in the Amsterdam

[12] https://web.archive.org/web/20181205153226/https://www.cbs.nl/nl-nl/dossier/nederland-regionaal/geografische%20data/wijk-en-buurtkaart-2017

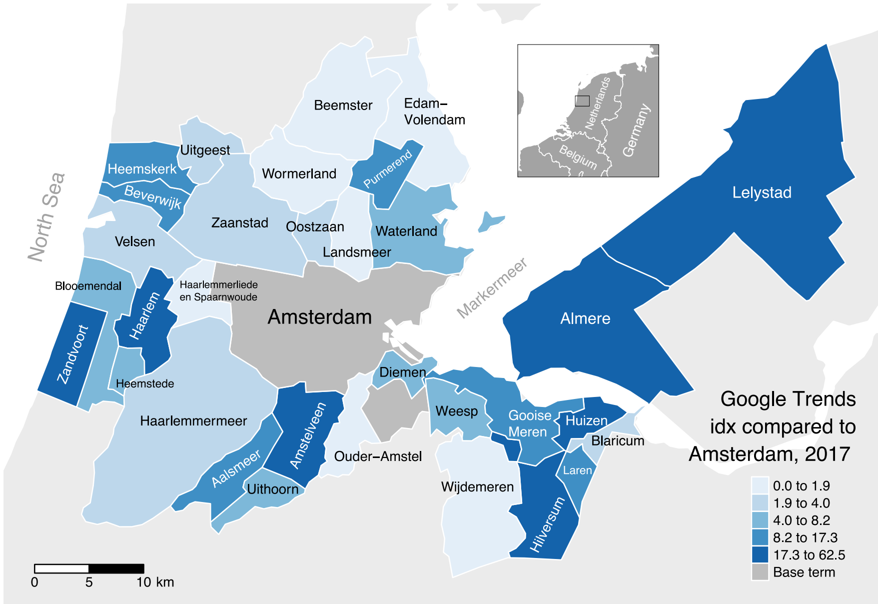

Fig. 4: Google Trends index for municipalities in the Amsterdam metropolitan region in 2017, using "Amsterdam" as the base term, from the whole world, without low-volume data. White areas represent water. Data grouped as quantiles.

area, and also have recreational highlights, such as a large shopping centre with modern architecture, a replica of a cargo ship of the Dutch East India Company, and one of the Factory Outlet Centres in the Netherlands.

As the population size of these municipalities varies from about 6,000 to 845,000, we related the GTI with population densities, by grouping population densities and GTI based on quantiles (RQ4). The bi-variate choropleth map in Figure 5 shows the geography of search interest in relation to population.[13] The brown municipalities (high population and high interest) dominate the metropolitan area, and are of particular interest for tourism. The light red areas manage to attract high interest with low population: Huizen, Laren, and Blaricum are gentrified neighborhoods, and a popular beach is located in Zandvoort. Haarlemmermeer and Velsen, in contrast, are much more affordable housing areas of lower interest.

By measuring the *demand* for an area in terms of search interest and its *capacity* in terms of population or infrastructure density, we suggest that the worldwide interest in the red and brown municipalities is likely to overcome their carrying capacity. Thus, rising numbers of residents and visitors can be expected with a further increase in property prices and the development as well as aggravation of capacity problems, including (over)crowding. This Google Trends data allows to identify, among others, Amsterdam, Almere, Lelystad, Harleem, Hilversum, and Hamstelveen as hotpots

---

[13] Map based on R package: https://github.com/sdesabbata/BivariateTMap

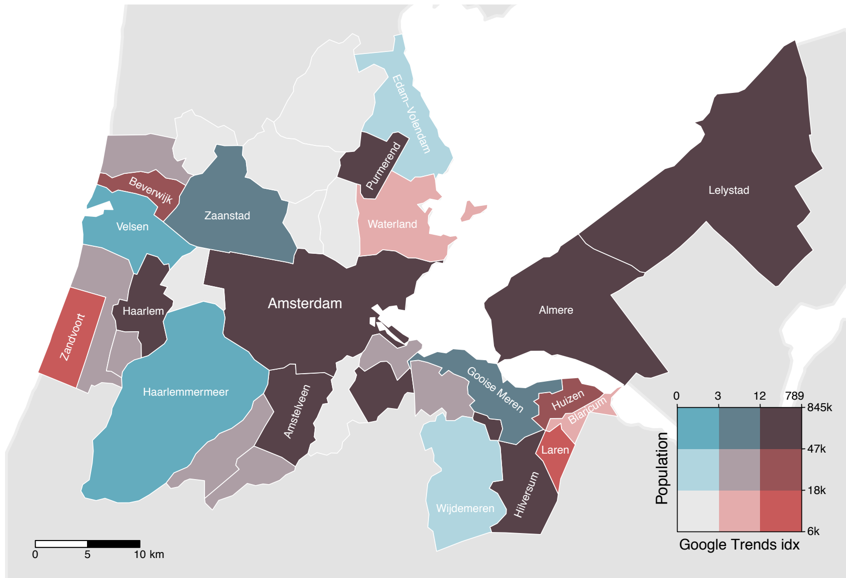

Fig. 5: Population of municipalities in the Amsterdam metropolitan region 2017, in relation to Google Trends index in period 2007–2017 (sum), without low-volume data. The 9 groups are calculated as the intersections of 3 quantiles for the two variables. For the sake of readability, labels are omitted for low-low and medium-medium areas. Sources: Statistics Netherlands and Google Trends.

that combine high population and high tourist in-flows, resulting in overcrowding (RQ4). By contrast, turquoise areas (high population and low interest) emerged as under-represented in the interest geography. These areas might be considered as targets for intervention to absorb tourist flows from high-interest areas (RQ4).

4.3 Interest within municipalities

To see further temporal and spatial trends in the Amsterdam metropolitan area (RQ1), we studied the temporal variation in GTI for each municipality separately, comparing only whether interest declined or increased in the last decade. This latter measure is comparable between regions, while the underlying absolute values are not. Rather than the variation with respect to a base term, in this instance, GTI is 100 at the peak of each municipality.

Figure 6 shows the change in GTI for each municipality, only considering searching originated from the Netherlands. The GTI increased for some municipalities (i.e. Amsterdam, Almere, Oostzaan, and Landsmeer, Edam-Volendam and Beemster, Blaricum and Laren), while other areas declined, including Haarlemmermeer, Zaanstad, and Wormerland. Possible explanations for the interest growth for some

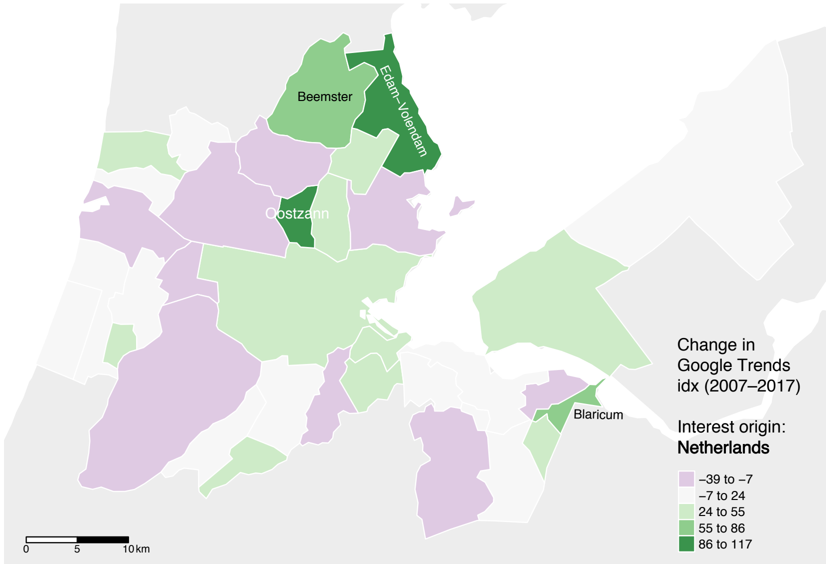

Fig. 6: Change in Google Trends index in each municipality, without low-volume data (2007–2017). Interesting origin: Netherlands. Data is grouped with equal intervals. For the sake of readability, some labels are omitted. Source: Google Trends.

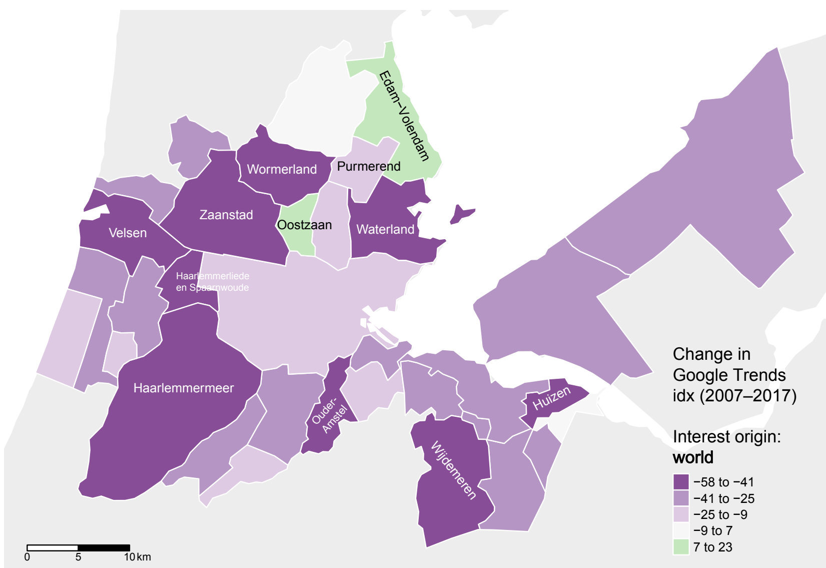

Fig. 7: Change in Google Trends index in each municipality, without low-volume data (2007–2017). Interesting origin: worldwide. Data is grouped with equal intervals. For the sake of readability, some labels are omitted. Source: Google Trends.

areas: Edam-Volendam, an old fishing village with a traditional cheese market, may have become more popular among visitors in recent years. Blaricum and Laren are among the most affluent municipalities in the Netherlands, attracting investors in real estate and luxury shoppers. Almere serves as one of the most important parts of the commuter belt of Amsterdam, and has absorbed some of the housing demand in the area, but also has appealing recreational facilities. Beemster has several historical fortresses that attract tourism. The popularity of Oostzaan may relate to the fact that it is located closely to the Zaanse Schans, a traditional working village with several windmills.

When observing global searches, the distribution of change looks similar, but with a steep decline over most regions (Figure 7). Even Amsterdam, Almere and Blaricum lost interest over the last decade, while only Beemster, Edam-Volendam and Oostzaan stayed equal or increased. To investigate the decline, it is beneficial to look at the variation in GTI at a higher temporal detail (Figure 8). Almost all areas declined in 2008–2011, with a particularly steep decline in the Western part of the region, with negligible exceptions. The decline is subsequently more biased towards the Eastern areas, with recovery localized in the North. Finally, from 2014 to 2017, more areas are either recovering or remaining stable.

This overall international decline in interest in the Amsterdam metropolitan area escapes simple explanations, and seems to counter other trends (RQ3). A possible factor can be identified in the sharp increase in mobile searches on smartphones, which increased and changed the search pool, therefore impacting on the GTI. While the 2008 financial crisis marked a temporary reduction in tourist flows, tourism at the global level seems rather "shock proof" and showing robust growth in the period 2007–2017. This suggests that spatially and temporally more detailed analyses are needed to relate this large decline to ground truths (RQ2 and RQ3).

5 Discussion and conclusion

In these analyses of Google Trends, we have shown several possibilities for empirical research. First, we studied the interest origin for the Netherlands and Amsterdam at the global level (RQ2), highlighting the rise of Europe as the main origin area, and how post-colonial ties are clearly visible in the spatial distribution of interest. Second, we compared search trends with hotel visits (RQ3). The correlation turned out to be fairly strong at the city level, but very weak at the country level (RQ1).

We then moved on to observe the interest geography at the municipality level, comparing areas to one another and in terms of population (*between* areas) and at the level of individual change (*within* areas). This analysis illustrated how uneven and volatile the geography of interest is, and how some areas appear largely over-represented in the interest they generated (e.g. affluent neighborhoods) and others are under-represented (e.g. residential areas) (RQ4). While this unevenness is perhaps not surprising, it is novel to view interest quantified over time at this scale (RQ1).

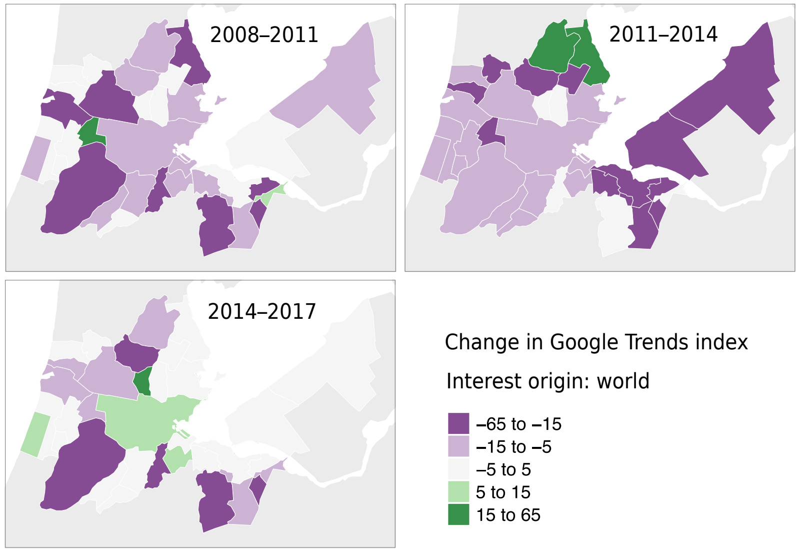

Fig. 8: Change in Google Trends index in each municipality, without low-volume data in 3 temporal intervals (2007–2017). Groups were manually defined. For the sake of readability, some labels are omitted. Interesting origin: world. Source: Google Trends.

For the case study of the Amsterdam metropolitan area, we have shown that it is possible to detect major areas of interest, in particular ones that attract a relatively high volume of searches despite low population density: These are potential hotspots of urban overcrowding. In addition, we might be able to detect suburban areas where interest is rising due to tourist attractions and appealing housing opportunities (globally or from a specific country of origin). These are areas where interest, by tourists as well as residents, is already rising. Processes of a rapid increase in interest, such as gentrification, might be detectable through this data. Furthermore, areas that are not yet on the tourist map could be identified for future marketing or development.

Rather than providing exhaustive answers to our research questions, this article opens up a path to further, more modelling-oriented investigations, adopting high-quality, tourism-related datasets as ground truth. While not reaching final conclusions, this exploration has provided us with several insights. In order to make effective use of Google Trends data in tourism studies, and more broadly, in geography, we deem the following points to be noteworthy:

- The geographic scale has a strong impact on the search behaviours for places. For example, patterns at the country and at the city level appear very different (e.g. city searches correlate to hotel bookings, country searches do not). Furthermore, for scales finer than a municipality, the data is not granular enough.
- The resolution of the GTI is intrinsically limited. Our algorithm (see Section 4.1) can help increase the resolution of the index resolution for relatively low-interest areas. However, peripheral small neighborhoods are likely to have a GTI $= 0$,

which cannot be re-scaled. Many prominent points of interest can attract higher interest than large spatial units (e.g. Van Gogh Museum).
- Semantic ambiguity of search terms is a critical problem that cannot be solved completely. The manual inspection of search results for each query in the study is highly recommended. An approach worth pursuing might be to distinguish between tourism, reference, and housing searches through more specific terms (e.g. "Amsterdam hotels" as opposed to "Amsterdam rent").
- Estimates from SEO companies such as SemRush can provide a quantification of search volume to complement the GTI, although the quality of such data is largely undefined.
- Searches for places tend to exhibit strong seasonality. Interest bursts are a serious issue and should be accounted for. Burst-causing events might be both unplanned (natural and man-made disasters) and planned (e.g. the World Cup).
- It is essential to possess local expert knowledge in order to interpret trends and spatial patterns both in interest origin and targets. Even so, in our case study, some patterns remained hard to interpret with obvious explanations (e.g. tourist flows). Place searches have their own peculiar, rather volatile geography, and some hard-to-explain variation must be expected.

This study of the spatial dimension of Google Trends data can enable a novel observation point on urban geographies. More efforts are needed to devise techniques to increase the semantic accuracy of terms, to reduce noise from bursts, and to disaggregate search behaviours that are currently conflated. Other kinds of ground-truth flow data (tourism and migration statistics, Airbnb stays, and smartphone data) could in the future be used to calibrate the Google Trends statistics. The inclusion of more granular search targets, such as a specific, highly visible points of interest, may also uncover meaningful patterns. Such big data sources might help gather new knowledge in the geographic domain, including the tourism-induced (over)crowding crisis that places increasing pressure on several cities around the world.

Acknowledgements The authors gratefully acknowledge Google for making some of its search data publicly available, Flavio Ponzio for providing insights on Search Engine Optimisation, and Stefano de Sabbata for his bi-variate choropleth library.

References

Askitas, N. and Zimmermann, K. F. (2015). The internet as a data source for advancement in social sciences. *International Journal of Manpower*, 36(1):2–12.

Ballatore, A., Graham, M., and Sen, S. (2017). Digital hegemonies: The localness of search engine results. *Annals of the American Association of Geographers*, 107(5):1194–1215.

Ballatore, A. and Jokar Arsanjani, J. (2018). Placing Wikimapia: an exploratory analysis. *International Journal of Geographical Information Science*, pages 1–18.

Ballatore, A., Scheider, S., and Lemmens, R. (2018). Patterns of Consumption and Connectedness in GIS Web Sources. In Mansourian, A., Pilesjö, P., Harrie, L., and van Lammeren, R., editors, *Geospatial Technologies for All. AGILE 2018*, pages 129–148, Berlin. Springer.

Choi, H. and Varian, H. (2012). Predicting the Present with Google Trends. *Economic Record*, 88(1):2–9.

Dergiades, T., Mavragani, E., and Pan, B. (2018). Google trends and tourists' arrivals: Emerging biases and proposed corrections. *Tourism Management*, 66:108 – 120.

Goel, S., Hofman, J. M., Lahaie, S., Pennock, D. M., and Watts, D. J. (2010). Predicting consumer behavior with Web search. *Proceedings of the National Academy of Sciences of the United States of America*, 107(41):17486–17490.

Gospodini, A. (2006). Portraying, classifying and understanding the emerging landscapes in the post-industrial city. *Cities*, 23(5):311–330.

Graham, M., De Sabbata, S., and Zook, M. A. (2015). Towards a study of information geographies:(im) mutable augmentations and a mapping of the geographies of information. *Geo: Geography and environment*, 2(1):88–105.

Hall, T. and Barrett, H. (2017). *Urban Geography*. Routledge, London, 5th edition.

Hendler, J., Shadbolt, N., Hall, W., Berners-Lee, T., and Weitzner, D. (2008). Web Science: An Interdisciplinary Approach to Understanding the Web. *Communications of the ACM*, 51(7):60–69.

Jun, S.-P., Yoo, H. S., and Choi, S. (2018). Ten years of research change using Google Trends: From the perspective of big data utilizations and applications. *Technological Forecasting and Social Change*, 130:69 – 87.

Kitchin, R. (2014). Big data, new epistemologies and paradigm shifts. *Big Data & Society*.

Lazer, D., Kennedy, R., King, G., and Vespignani, A. (2014). The Parable of Google Flu: Traps in Big Data Analysis. *Science*, 343(6176):1203–1205.

Li, X., Pan, B., Law, R., and Huang, X. (2017). Forecasting tourism demand with composite search index. *Tourism management*, 59:57–66.

Neuts, B. and Vanneste, D. (2018). Contextual Effects on Crowding Perception: An Analysis of Antwerp and Amsterdam. *Tijdschrift voor economische en sociale geografie*, 109(3):402–419.

Novy, J. and Colomb, C. (2017). *Protest and Resistance in the Tourist City*. Routledge, London.

Önder, I. and Gunter, U. (2016). Forecasting tourism demand with Google Trends for a major European city destination. *Tourism Analysis*, 21(2-3):203–220.

Pan, B., Chenguang Wu, D., and Song, H. (2012). Forecasting hotel room demand using search engine data. *Journal of Hospitality and Tourism Technology*, 3(3):196–210.

Popp, M. (2012). Positive and negative urban tourist crowding: Florence, Italy. *Tourism Geographies*, 14(1):50–72.

Siliverstovs, B. and Wochner, D. S. (2018). Google trends and reality: Do the proportions match?: Appraising the informational value of online search behavior: Evidence from swiss tourism regions. *Journal of Economic Behavior & Organization*, 145:1–23.

Singleton, A. D., Spielman, S., and Brunsdon, C. (2016). Establishing a Framework for Open Geographic Information Science. *International Journal of Geographical Information Science*, 30(8):1507–1521.

Spierings, B. (2013). Fixing missing links in shopping routes: Reflections on intra-urban borders and city centre redevelopment in Nijmegen, The Netherlands. *Cities*, 34:44–51.

Stephens-Davidowitz, S. I. (2013). *Essays using Google Data*. PhD thesis, Harvard University, Cambridge, MA.

Yang, X., Pan, B., Evans, J. A., and Lv, B. (2015). Forecasting Chinese tourist volume with search engine data. *Tourism Management*, 46:386–397.