# AN EXPLORATION OF INTERACTIONS BETWEEN URBAN HERITAGES AND TOURIST'S DIGITAL FOOTPRINT: NETWORK AND TEXTUAL ANALYSIS VIA GEOTAGGED FLICKR DATA IN AMSTERDAM

S. S. Karayazi<sup>1, \*</sup>, G. Dane<sup>1</sup>, T. Arentze<sup>1</sup>

<sup>1</sup>Urban Systems and Real Estate, Department of Built Environment, Eindhoven University of Technology, The Netherlands (s.s.karayazi, g.z.dane, t.a.arentze)@tue.nl

#### Commission IV, WG IV/9

KEY WORDS: LBSN data, Flickr, Network mining, Text mining, Heatmap, Amsterdam, Urban heritage, Tourism

#### **ABSTRACT:**

Spatial information retrieved by geotagged social media data allows the analysis of when and where questions and supports a better understanding of people's spatial interaction. The purpose of this paper is to reveal alternative heritage spots for reducing tourism pressure in certain areas within Amsterdam's historical core by presenting tourists' points of interests (POIs), their patterns between POIs, and the sentiments attached to these POIs. To fulfil the research aim, network analysis, textual analysis, and heat maps were conducted with geotagged Flickr data representing tourists' uploaded images. The results show that the Flickr data set was useful to reveal the existing POIs and their relations. Moreover, the further analysis that is based on the intersection between the revealed spatial patterns and Amsterdam heritage sites showed alternative POIs nearby current POIs. Based on the results, we demonstrated how a dataset constructed from geotagged Flickr data can provide useful practical information for sustainable tourism development. Our research has the potential to support urban heritage and tourism researchers and policy makers with a better understanding of POIs, their relations and possible other POIs. The outcomes of this research can advise heritage tourists on alternative POIs to visit, and policy makers to plan alternative POIs.

#### 1. INTRODUCTION

Over the last years, increasing tourism activities and the amount of tourists in European cities such as Amsterdam, Barcelona, Venice, Berlin, Copenhagen, have been detrimental to heritage artifacts and their values that are considered as magnets for tourists (Hospers, 2019). According to The World Tourism Organization (UNWTO, 2018), overtourism is described as "the overcrowding from an excess of tourists and its impact on a destination, or parts thereof, that excessively influences perceived quality of life of citizens and/or quality of tourist experiences in a negative way." In that sense, destinations that are attractive to tourists are affected by the consequences of mass tourism activities. Therefore, an increasing need emerges to understand and monitor tourist patterns by considering their relation with the environment in space and time in order to combat the challenges of overtourism.

In historic cities, certain heritage buildings and areas become very attractive to tourists because of their significance, overpromotion, and popularity (García-Hernández et al., 2017), while some heritage areas may not attract tourists and become underrepresented. There is a need to understand tourists' movements in the city, and how tourists perceive and experience different places (Trinh & Ryan, 2017). Based on this understanding, tourists can be provided with location suggestions not only of well-known places but also underrepresented areas with similar heritage types for reducing tourism pressure by distributing tourists throughout the city. That can lead to developing sustainable tourism policies in overly touristic heritage cities.

With the emergence of big data from newly available sources, such as location-based social networks (LBSN), volunteered geographic information (VGI), such as Flickr, Foursquare, which

provide a wide range of spatial and non-spatial information, it is possible to conduct evidence-based urban studies research (Song & Liu, 2017). The dimensions of urban tourism and heritage studies that utilized LBSN data sources differ. For instance, van der Zee et al. (2018), focused on destination management by utilizing the TripAdvisor dataset to reveal tourist's digital footprint from five Flemish historical cities. They found that spatial clustering and hotspot analysis are capable of showing tourists' patterns and providing new knowledge about actual use of space by users and its online representation. Karayazi et al. (2021), utilized the combination of multisource information such as heritage data, supporting products data (i.e., attractions, museum, open market, shopping) and Flickr data to understand attractive heritage locations and the built environment characteristics that make them attractive. They conducted cluster analysis with density-based spatial clustering of applications with noise (DBSCAN) to identify POIs, then employed ordinary least square and geographically weighted regression, and revealed that tourist's POIs were concentrated within Amsterdam's historical core whereas locals were distributed over the city. In addition, they emphasized that the combination of less attractive heritage with strong influential supporting products (i.e., tram-metro stations) could introduce to sustainable tourism by promoting less attractive heritages transforming into attractive in Amsterdam. Ginzarly et al. (2018), conducted an analysis of Flickr photos and the analysis of their tags for mapping historic urban landscape. First, they classified photographs as tangible and intangible images to understand what the scenes depict. Second, they focused on textual analysis of Flickr tags with quantitative and categorical analysis. They found that the tags could explain the spatial perception of users. These studies concluded that LBSN data is qualified to capture the relations between visitors' movements and perceptions. Regarding the methodology to utilize newly available data such as Global Positioning System (GPS) and LBSN, in tourism literature,

<sup>\*</sup> Corresponding author

several clustering algorithms and network analysis have been frequently applied for understanding tourist movements and patterns. Dane et al. (2020), focused on visitor flows utilizing GPS data to detect the area of interest locations. They analysed visitors' temporal and spatial behaviour including origindestination and intra-event destinations by means of network analysis. Grinberger et al. (2014), identified a group of tourists behaviour with GPS trackers using a clustering algorithm. Hu et al. (2019), detected the tourist movement patterns by combining three methods. First, they collected georeferenced Twitter data published by tourists. Then, the authors used the DBSCAN algorithm to find tourist clusters. Finally, they applied a network analysis algorithm including in-degree, out-degree, and betweenness degree to find popular attractions, centric attractions, and popular point-to-point routes. Although LBSN data sources provide location and time information, they are not generated for the purpose of analysing human spatial behaviour due to the nature of data. There is a need to understand what meanings are attached to spaces by tourists. Hauthal & Burghardt (2016), employed natural language processing (NLP) to words that were extracted from the title, description, and tag of the Flickr photographs to detect the activities and emotions of locals and tourists. This study demonstrated tourists' emotional activity and their perceptions with LBSN data via text mining including sentimental and lexical analysis. Such information can provide insights into how tourists interpret the spaces.

However, the viewpoint of these above-mentioned studies that utilize newly available datasets, is either about tourist behaviour without a focus on heritage significance or about heritage as a historically significant place within the urban level and missing the tourism perspective. To bridge the gap, we combine both perspectives with tourist data from Flickr and building-based heritage data from Amsterdam. By building on previous research, our paper presents an approach to understand spatial interactions between urban heritages and tourist patterns in Amsterdam with Flickr data and heritage data from Amsterdam for supporting policy makers to reduce tourist pressure in certain areas. To do that, this study follows three steps: (i) to define tourists' patterns between the found POIs via network analysis, (ii) to define how tourists perceive the POIs by using textual analysis that utilizes image captions from Flickr database so that the relation between these POIs and the meaning attached to them is captured and (iii) to investigate the intersection between the existing heritages and the Flickr LBSN data to reveal the alternative POIs for future as alternative to mostly visited POIs.

The remainder of the paper is organized as follows: Materials and methods section explains the study area and data collections and methods that were used in this research. Next, the results of each step are presented. Finally, conclusions are given with the discussion of findings, the suggestions for policymakers, and a discussion of the limitations of the current study.

# 2. MATERIALS AND METHODS

# 2.1 Study area and data characteristics

In this research, Amsterdam was chosen as a case area where 45.9 million tourists were counted in 2019 and 27.3 million were counted in 2020 (CBS, 2021). Although, the number of tourists was reduced substantially due to restrictions in 2020, pre-Covid-19 projection was 30 million tourists in 2025 (Hospers, 2019). Therefore, Amsterdam's city council initiated the "City in Balance" program that proposes to find a balance between tourists and locals. The city council seeks alternative destinations

to expand Amsterdam's tourist concentration and they aim to promote other less visited places. From an urban heritage perspective, Amsterdam is home to unique historical buildings which are protected by United Nations Educational, Scientific and Cultural Organization (Centre, 2020), and over-touristic activities might have a destructive effect for Amsterdam's heritages. To prevent the influence of mass tourism, there is a need to reveal alternative intra-destinations (POIs) within Amsterdam based on an understanding of the relations between tourists' behaviour and heritages.

We used two different datasets for this current research. First, Flickr was utilized to detect the POIs for tourists which refer to the clusters of a high number of photographs taken and tagged at these locations and to understand the meanings attached to each POI by using the tags and descriptions of uploaded Flickr images. The main contribution of Flickr photographs is that these images represent attractive places with geotags. Second, we used National Monument dataset from the Cultural Heritage Agency website of the Dutch Ministry of Education, Culture and Science (Ministerie van Onderwijs, 2019) for analysing the relations between heritages and found POIs in order to detect alternative POIs.

The metadata of photographs were downloaded following these coordinates; "minx": 4.867080, "miny": 52.357924, "maxx": 4.933176 and "maxy": 52.390259 which represent the boundary of central Amsterdam (Figure 1). Flickr API offers metadata of photograph's and in this study, the photograph's location, unique owner ID, photograph's taken time and photograph's textual data (except URL) were used. User-related information was removed due to protecting users' privacy. The Flickr dataset was cleaned by removing the duplicate and invalid records to minimize the dataset's errors. The timestamps of downloaded photographs were used to divide the users into two groups, namely tourists and locals, by considering photographs' taken time range. As suggested in the literature (Girardin et al., 2009; Hauthal & Burghardt, 2016; Karayazi et al. 2021., Koutras et al., 2019), we separated the taken times of photographs into 30 days period. To illustrate, if users uploaded more than one photograph within the assigned consecutive 30 days, the algorithm assigned the user as a tourist. If the duration between the uploads of two photographs for a given city is greater than 30 days, the user was assigned as a local.



Figure 1. Flickr photographs (Karayazi et al., 2021)

Figure 1. represents the distribution of tourists' and local photographs. The raw dataset contained 285.130 photographs which were taken between 1927 and 2019. We checked

photographs that were taken around 1927, and we kept these photographs in the dataset since they have still represented spatiotemporal information. After cleaning, 12.766 photographs remained from 1808 tourists, and 25.445 photographs remained from 654 locals. Note that local's photographs will not be analysed further, because this research focuses on spatial interactions between urban heritage and tourist patterns in Amsterdam. A national monuments dataset of Amsterdam (Ministerie van Onderwijs, 2019) was used to understand relations between tourists and urban heritages. The dataset is presented in shapefile format and contains 7500 registered heritages in Amsterdam. Each heritage has its attributes such as identity, coordinates, function, and postcode. For this research, 25 predefined heritage types were aggregated to 3 groups such as culture and education (i.e., church, culture-sport, education), commercial and governmental (i.e., governmental building, industry, shopping, storage, transportation, office building), and recreation (i.e., catering, garden and park, zoo), as explained in Karayazi et al. (2021) to avoid spatial bias. The reason was that some of the heritage types had less representation; therefore, it could cause misinterpretation for further analysis.

# 2.2 Methods

In this part, we elaborate the data preparation process and steps. For the study, Gephi 0.9.2 version for network analysis, ArcGIS Desktop 10.8 version for visualization and generating heatmaps, Microsoft Azure Machine Learning add-in for text sentiment analysis, and online word-cloud generator (Zygomatic, 2016) to observe lexical differences were applied, respectively.

2.2.1 DBSCAN and Network Analysis: In this study, we used the clusters from Karayazi et al. (2021) to conduct the network analysis. Each cluster represents the POIs, and these concentrated locations can be defined as the most photographed areas by the tourists. DBSCAN is spatial clustering methods which transforms an object into clusters that have similar specifications within groups that can be defined as high homogeneity. On the other hand, dissimilar qualifications among the other groups can be accepted as high heterogeneity. DBSCAN is widely used in urban planning studies with big data (Devkota et al., 2019; Karayazi et al., 2021; Koutras et al., 2019). The algorithm searches for high-density areas with two parameters; areas of the neighbourhood (eps) and minimum points (MinPts) within these areas. The resulting clusters are referred to as POI which defines the nodes for the network analysis.

Network analysis that is derived from graph theory seeks to explain the structure of relations between given objects (Shih, 2006). Network analysis in tourism literature was developed under three research themes: collaboration and knowledge creation; tourism supply, destination and policy systems; tourist movement and tourist patterns (Liu et al., 2017). In that sense, we focus on network analysis for identifying tourist patterns via geotagged Flickr data. In this study, network analysis is based on nodes (i.e., vertex) which are the POIs that were extracted from DBSCAN, and links (i.e., edge) that define the tourists' movements between these POIs. The sequence of photographs where tourists move from one place to another can reveal their visitation preferences. Therefore, two different matrices were designed namely nodes and edges. Nodes represented the origins and the destinations of the trips, and edges represented the links that were based on tourists' movement between the nodes. The network of nodes was examined as directed and weighted. From Flickr data, we retrieved temporal information of each photograph within the POIs, then organized them considering the

time sequences. In this research, if tourists take one photograph within the POI, and then take another photograph within another POI it is assumed that there is a connection between two POIs.

In this current study, network analysis was used to analyse the spatial centrality of the most attractive POIs and the spatial interactions between them. In the literature, the most common types of centrality, which reveals the most remarkable nodes that are located at the crucial locations within the network, are identified based on degree (highest degree means the highest centrality), betweenness (highest betweenness indicates significant intermediary), closeness (shortest path among the nodes) and eigenvector centrality (the importance of connection) (Dane et al., 2020; Shih, 2006). Degree centrality addresses the importance of the given nodes within the graph, and it explains the number of direct connections each node has to the other nodes in the network. It calculates a score among the links associated with each node. The higher degree of a node, the more crucial it is in a network. Betweenness centrality measures the number of times a node functions as a connection along the shortest path between a given node (Saxena & Iyengar, 2020). It is calculated by classifying all the shortest paths between all pairs of nodes and measuring how many times each node occurs on these paths. For this research, it revealed the important POIs that act as a bridge for tourist flows. Closeness centrality measures the closeness of a given node with other nodes of the network. The highest score represents the highest influential node which has a shorter distance to other nodes. Therefore, it was interpreted as a reachable POI because it plays a pivotal role within the network. Lastly, eigenvector centrality measures the level of a node's influences within a network. It is calculated by evaluating how a single node is well or ill-connected throughout the network with the greatest connectivity. Nodes with high eigenvector centrality scores might have multi-connections and their connections might have other connections (Bihari & Pandia, 2015). For this research, it was used to understand the highly influential POIs and these POIs' connections.

**2.2.2 Text analysis:** Our research includes textual analysis such as sentiment analysis to reveal space-related emotions, and text mining to understand space-related interpretation. For text mining, tags and descriptions of each Flickr photograph were utilized. The aim of textual analysis was to reveal the sentimental and lexical differences of texts (i.e., tags, descriptions) extracted from Flickr data. Azure Machine Learning is based on an NLP algorithm and a built-in dictionary with positive and negative words. Each word in the glossary is defined by the negative, positive, or neutral value. In our study, sentiment and score in each photograph were assigned based on the corresponding text. For the lexical differences, the most used words were analysed with online word cloud generator. Highlighted words were appeared bigger than less used words.

**2.2.3 Heatmap:** Heatmaps are a graphical representation of spatial data that consists of a set of cells where the special values attributed to the cells are shown. The "heat" can be described as a high concentration of geographical objects in a specific place, and they can show surface density that visualizes the location of concentration points. The Flickr photographs and heritage artefacts were analysed by applying the Kernel Density algorithm. It calculates magnitude-per-unit area from point or polyline features using a kernel function to fit a smoothly tapered surface to each point or polyline. As our study is based on point features, we generated the density of linear features in the neighbourhood of each output raster cell (Kernel Density, 2022). In this research, heatmaps are used to show the heritage density to visualize the interaction between the heritage artefacts and

POIs. Based on that understanding, suggestions can be provided to tourists for other potential heritage sites that can be interesting to them. A heatmap of the location of Flickr photos taken by tourists and the multiplication of it with the heatmap of heritage data considering their types, can be created. This understanding can be used to give suggestions toward less densely areas. Raster multiplication allows observing where overlay(s) two raster in space. To conduct multiplication, the map algebra function was used in the ArcGIS environment. The expression of the function was generated by specifying inputs, values, and operators. In this section, we used Flickr data and heritage data considering their type as inputs, multiplication (\*) as an operator. As a result, an output raster was generated that visualizes the surface density of Flickr and heritage data.

With the following three-step approach, we present tourists' patterns and their interpretations, and aim to reveal alternative POIs with heritage data for reducing tourism pressure within certain areas in Amsterdam's historical core.

#### 3. RESULTS

The spatial distribution of Flickr photographs on heatmap is given in Figure 2. The raster values were classified with natural breaks function using 10 classes. Note that the value of 0 has been assigned as transparent for a better interpretable base map.



Figure 2. Heatmap of tourist's photographs

In Figure 2, red areas represent highly photographed areas based on Flickr photo uploads of tourists. On the contrary, blue areas represent lower densities. Tourists took photographs mainly around Amsterdam Central Station and Museumplein as was expected due to these locations' popularity. In addition, places outside Amsterdam's historical core such as Het Schip, Sloterdijk, and NDSM Werf were also photographed. This heatmap helps to identify not only the places with the largest concentration of tourists but also determining the places where the tourists are represented in less quantities which are valuable for the analysis in this paper.

#### 3.1 DBSCAN and Network result

As was indicated in section 2.2.1., we used clusters from Karayazi et al. (2021) as input in the network analysis. The authors identified 9 POIs for tourists (Figure 3) using the following parameter settings: minPts=125 and eps=70. In Figure 3, it is seen that the larger cluster represents the larger quantity of photographs. While the most photographed area was

Museumplein (T2), the least photographed area was Het Schip (T9). The results also show consistency with the heatmap presented in Figure 2.



Figure 3. Distribution and number of photographs per POIs

To better understand the relations between these clusters that represent POIs, network analysis was conducted. From here on, POIs will be named with their T codes. The result of centrality indicators for the network analysis can be seen in Table 1. T1 has the highest degree, and it is followed by T2, T7 and T6. Considering the closeness centrality (CS) T1 and T2 have the same value, and they are followed by T7, T4, T6, T5, and T8. However, T9 is calculated as 0 which means that it was not as easily reachable as other nodes. For the betweenness centrality (BC) T1 has the highest value, and next T2 and T6 respectively. It means that people who took photos at T1, T2 and T6 also took photos at other POIs. For the eigenvector centrality (EC) T1 is equal to 1 which indicates that it is connected with other POIs very well. Among the other POIs, T2, T6, and T7 have the same value and are closest to 1, i.e., they are approximately 0.94.

ID	Degree	CS	BC	EC
T1	16	0.889	11.583	1
T2	15	0.889	9.083	0.937
Т3	5	0.500	0	0.377
T4	10	0.727	0.250	0.557
T5	10	0.667	0	0.738
T6	13	0.727	1.083	0.937
T7	14	0.800	4.000	0.937
T8	10	0.667	0	0.666
T9	3	0	0	0.184

Table 1. Centrality indicators

The visual result of the network analysis can be seen in Figure 4. The size of the POIs is equivalent to the degree of centrality measurement. The colour of POIs is assigned based on the betweenness centrality measurement (darker node means higher betweenness centrality), and the colour of the links is assigned based on the weight of the links (darker link means higher weight).

As it can be seen in Figure 4, T1 and T2 were found to be well connected. Therefore, after these POIs, tourists were distributed within the inner city of Amsterdam. This might mean that T1 and T2 were the main attraction and distribution points. Another

strong connection was found between T1 and T6. Due to the location, the church is easily accessible, and it is well-connected to other nodes such as T7, T2, and T8.



Rijksmuseum. Other artistic words such as "night watch, Goya, Banksy, whistler" can be seen in Figure 5. (b).



Figure 4. Network analysis

However, T9 was only connected with T2. The reason might be that T9 is a museum, and it could appeal to a special target group, not all the tourists. These results reveal that alternative POIs that will be suggested should be reachable (or in close proximity) from T1, T2, and T7, T6 and T4.

## 3.2 Textual analysis result

The textual analysis aimed to reveal the sentimental and lexical differences of texts (i.e., tags, descriptions) extracted from Flickr photographs taken and uploaded by tourists and the 10 most frequent words for each POI. With the findings, we can evaluate whether found POIs represent tourism and/or heritage values.

The sentimental analysis resulted in 96% positive, 4% neutral, and 0% negative expressions. This result indicates that Flickr users do not upload photos that are associated with negative sentiments. Therefore, sentiment analysis will not be processed further. In this section, word clouds with the 10 most frequent words for each POI and their meanings is explained. Overview of the word cloud can be seen in Figure 5. T1 is located in the northern part of Amsterdam and can be seen as the gateway of the city. The station has direct connections to Schiphol Airport and large-scale cities in Germany, Belgium, and France. In addition, passengers can access other means of public transport such as underground, tram, and bus. According to the results, the 3 most used words are "Amsterdam", "Netherlands", and "Noord Holland". The reason might be that the station is the major arrival spot for tourists, and they narrated the Netherlands in the description of their photographs. In Figure 5. (a), "pays bas" which stands for the Netherlands in French was found 8th most used word. T2, Museumplein (museum square) is situated in the southern part of Amsterdam and is home to Rijkmuseum, Van Gogh Museum, and Stedelijk Museum. T2 also accommodates events, festivals, and celebrations. As it is expected, the most used words were "Rijkmuseum, museum, Netherlands". Also, "museo" which stands for the museum in Spanish was found 7th most used word. In this POI, tourists added tags to the photographs related to an artistic value such as "Rembrandt" is a famous Dutch painter whose art works are exhibited in

Figure 5. Word clouds with the 10 most frequently words

T3 is located southeast part and outside the historical core of Amsterdam. The most used tags are "ecir, Europe, qseu, Amsterdam, quantifiedself". We searched the meaning of these words on the Internet database, and it was found that "quantified self and qseu" belonged to a theme of a conference and meeting series that were placed in Hotel Casa (Quantified Self, 2022). Moreover, "ecir" is a conference that represents European Conference on Information Retrieval and was organized in Hotel Casa in 2014, therefore "europe" might be associated with conference name and attendees might add region's information to their photographs (ECIR, 2014). These POI's tags do not represent any heritage value considering location itself and tags and descriptions of the Flickr photos. T4 is a film museum situated at the north bank of the IJ river, opposite to the Amsterdam Centraal. The most used tags were "Amsterdam, lookouts, eye, Netherlands and a'dam". "Lookout" represents the A'dam Lookout which is an observation deck located on top of the Amsterdam Tower. In this POI, tourists added tags to the photographs regarding two groups; Eye Film Museum and A'dam Lookout which are located approximately 250 meters from each other. Other most-used tags were "architecture, river and xxx". "Architecture" might depict the exterior of Eye Film Museum due to the outstanding design and "xxx" represents the city's coat of arms. T5 consists of a registered monument, De Oude Kerk, and serves as a museum, church, and contemporary art institution which is in the historical core of Amsterdam. The most used tags were found related to church and church's architectural descriptions such as "gothique, blazon, clocher". T6 is located on the opposite side of the Amsterdam Centraal Station. St. Nicholas Church serves as a museum church and tourists can benefit from the guided tour. Therefore, most used words were found related to function such as "church, Nicholas, basilica". In addition, users add "water, canal" due to the location, "architecture, heritage" to explain the Church's architectural and monumental value. T7 is not only a well-known and popular landmark but also a public space that is a gateway to Madame Tussauds Museum, Nieuwe Kerk, and Royal Palace in Amsterdam. Textual information about the Dam was found "koninklijk palaise Amsterdam, street, noiseless". In addition, "urban, city life, tourist" were found in the tags of the photographs. T8 is a historic brewery museum located close to

the Museumplein. The brewery is an anchor point on the European Route of Industrial Heritage which presents more than 300 members in 30 European countries such as Netherlands, Belgium, and Chezhia (ERIH, 2022). The most common words found were "Heineken, bottle, experience, green". T9 is a museum that is known as The Ship due to the exterior design of the building. The Ship is located top of the Amsterdam School which can be described as a movement that has an influence on the construction and decoration over the Netherlands (Het Schip, 2010). The building is used for temporary exhibitions and museums. The most used words were "Amsterdam, amsterdamse school, het schip, architecture". In addition, "michel de klerk" and "de klerk" represent the architect's name are added frequently to the photographs.

The extracted most used words derived from tags and descriptions were useful to grasp how people interpret and what the attached meanings to the POIs are. Based on the results, tourists preferred to add where the photographs were taken as country and province name (i.e., Netherlands, Europe, Noord Holland), what the use of the photographed artefacts (i.e., museum, station), and the significant elements in photographed artefacts (i.e., vitraux, glass,), and the significant values of the photographed areas are (i.e., architecture, Rembrandt). Except T3, all other defined POIs were described as connected to heritage and/or tourism related use, elements and values. This indicates that Flickr data was useful to identify tourism POIs of Amsterdam.

#### 3.3 Heatmap of Heritage Artefacts

To understand the intersection between the photographed areas by tourists (including the places other than extracted POIs, as shown in heatmap Figure 2) and the type of heritage artefacts, we performed a series of raster multiplications. To do so, we multiplied the value of two raster's for instance a heatmap of tourists' photographs \* culture-education heritages on a cell-bycell basis. The results provide a better understanding of the spatial intersection of tourists' photographs and heritages with the heatmap, and they can reveal alternative POIs. The heatmaps of heritages per type and multiplication with tourists' photographs can be seen in Figure 6. The values of the raster were classified with natural breaks function using 10 classes.

According to the results, culture and education heritages were concentrated in Amsterdam's core (Figure 6.a). On the other hand, the multiplication heat map in Figure 6.(b) shows alternative POIs outside the core. Also, considering the network analysis (section 3.1), other location suggestions can be given to offer new alternative POIs to be promoted for tourists' visitations. For instance, users who took photographs around T1, T2, T6, and T7 might be interested in a visit to Jordaan, Haarlemmerbuurt, and Frederik Hendrickbuurt (yellow in Figure 7) because these neighbourhoods were found relatively attractive, and there exists culture and education heritages. Commercial-governmental heritages are densely located on the northwest and southeast axes, and the south part of Amsterdam (Figure 6.c). The multiplication map in Figure 6. (d)., had followed the same pattern. Zeeburg, De Pijp, and Transvaalbuurt (red in Figure 7) might be alternatives for tourists due to high ranking in correspondent rasters. Although these locations were not considered as POIs according to the results of the DBSCAN, they were found relatively attractive according to the heatmaps of tourists' Flickr photographs. Recreational heritage locations (Figure 6.e) are located more scattered than culture education and commercial-governmental heritages. The heritages are concentrated around the northwest and southeast axes. In addition, NDSM Plein (green in Figure 7) is rich in terms of recreational heritages.



Figure 6. a. culture-education heritages b. tourists' photograph\*culture-education heritages c. commercialgovernmental heritages d. tourists' photograph\*commercialgovernmental heritages e. recreational heritages f. tourists' photographs\*recreational heritages

The multiplication map in Figure 6. (f). reveals that tourists around T9, might be advised to visit around NDSM Plein which is located just outside the historical core of Amsterdam and reachable from T1 by boat and T4 by boat and on foot.





It should be noted that these suggestions can be extended to other heritage artefacts and areas in Amsterdam, via an in-depth analysis of these maps together with urban planning, tourism and heritage experts of Amsterdam.

## 4. DISCUSSION AND CONCLUSION

A three-step approach including cluster analysis and network analysis, text mining and heatmaps and raster multiplication analysis are introduced to understand spatial interactions between urban heritage and tourist patterns in Amsterdam via locationbased social network data and heritage data for supporting policymakers to reduce tourist pressure in certain areas.

In this study, the first research question is aimed to define tourists' patterns between the found POIs from Karayazi et al. (2021) via network analysis. We compared these POIs (Figure 3) with TripAdvisor's "Things To Do" list within Amsterdam using the search bar, and our results were compatible with TripAdvisor's database (TripAdvisor, 2022). For instance, Centraal Station is ranked 6 of 916, and Museumplein is placed 28 of 916 in the POI and landmark list. In that sense, Flickr data and the application of DBSCAN to this data can be considered useful to determine POIs. For the network analysis, Centraal Station and Museumplein were found well connected and strongly tied together. Considering the location of these POIs, they are well connected to each other not only by foot; they are also reachable by public transport. Most tourists may start their trips from Centraal Station, and they may follow directly to Museumplein which is home to the well-known Rijks Museum and Van Gogh Museum. However, in the data, the Het Schip only connected with Museumplein. The reason might be that Het Schip is a museum, and it could appeal to a special target group, not to all the tourists. Considering the degree centrality among the nodes, Centraal Station, Museumplein, and Dam Square have the highest rankings and it makes these POIs important.

Following that we investigated how tourists perceive the POIs, by using textual analysis that utilizes image captions and tags from the Flickr database. According to the results, sentimental differences cannot be observed. The reason might be that Flickr is a photograph-sharing platform at which users possibly prefer to upload photographs of positively experienced places. For the lexical distribution, users generally added the place name of the photographs taken such as Amsterdam, the Netherlands, and what photographs displayed such as canal, sculpture. Also, another finding is that words from several languages (except English and Dutch) appeared in the word clouds. It is an indicator that presented POIs appeal to international tourists. Although Flickr does not have any limitation to adding tags and descriptions, users did not prefer to attach too much textual explanation. The reason might be that Flickr is not meant for social interaction, users add tags and descriptions considering the best explicatory words so that they may draw other photographers' attention. Lastly, heritage-related words were observed, especially with respect to use and value. In that sense, users choose generic words instead of focusing on an in-depth meaning of the place. Overall, the findings of textual analysis indicate that Flickr data was useful to identify tourism and/or heritage POIs of Amsterdam.

The third research question attempted to explain the intersection between the existing heritages and the tourists' photographs based on heatmap analysis to reveal alternative POIs. The contribution of this question is to display other possible alternative locations to distribute tourists in space. According to the results, alternative POIs were suggested considering the heritage types. Regarding the heritages within the POIs boundary, a total number of 24 heritages exists out of which 10 were for cultural-educational, 11 were for commercialgovernmental and 3 were for recreational heritages. According to the results of the network analysis, T1, T2, T6 and T7 were found to have the highest degree of centrality indicators. Besides, T2, T5, T6 and T7 contains mainly cultural-educational heritages. T6, T1, T5, T7 and T8 contain commercial-governmental heritages, and they represented focal destination such as origin, core due to highest centrality. Therefore, commercialgovernmental heritage tourists might be guided starting from T1. In addition, the north bank of the IJ river might be visited by recreational heritage tourists. According to the results, recreational heritages are placed within T8 and T1. While T1 was found to have the highest value of betweenness centrality, T8 was found 0. It means that T8 does not act as significant intermediary node that bridge the interaction. Therefore, locations that are close and well-connected to T1 might offer better alternative locations. To do so, tourists may be distributed evenly outside the Amsterdam's historical core by suggesting existing heritages that were photographed but not found highly attractive. Based on these analysis steps, new alternative POIs could be suggested to tourists according to their interest on heritage types (see Figure 7). This study provides an evidence-based approach to defining alternative POIs for tourists in order to distribute them in space to overcome overtourism and its effects. Finally, these suggestions can be extended to other heritage artefacts and areas in Amsterdam, via an in-depth analysis of these maps together with urban planning, tourism and heritage experts of Amsterdam.

Overall, the 3-step methodology presented in this study provides a replicable approach based on Flickr LBSN and heritage data. This study shows that visualizing and analysing tourists' patterns with their digital footprint provides useful insights because tourists are both producers and consumers of such data. Moreover, irreplaceable heritages can be protected from possible tourist overload and other less known but important heritages can be promoted.

In this study, the presented results are promising; however, several limitations are identified. First, we utilized LBSN data from Flickr and users who are not familiar with Flickr may be skewed in the analysis. Additional LBSN data such as Foursquare, TripAdvisor, and Twitter may improve the accuracy. Further research can be conducted for cross-validation by making a comparison with these datasets. Grimaldi et al. (2021) revealed that corresponding restaurant's reviews from TripAdvisor reflect the inspection rating, therefore such real-time data could be useful to prevent infection among citizens. In that sense, online textual data from Trip Advisor can be used for further interpretation about spatial interaction between urban heritages and tourists. Second, we designed matrices for network mining considering an aggregated approach in terms of temporality. We did not consider the temporal distribution day by day; instead, we focused on the yearly distribution. If further studies include daily distribution with a sequence of photographs taken, it might provide an indication of trip chains. For example, assuming that a user who took photographs starting from Amsterdam Centraal and continued to the Museumplein and upload them respectively based on location sequence. However, a problem is that a user may visit another place between 1st and 2nd location and may not upload photo(s) to Flickr. On the other hand, seasonal differences in terms of temperature, weather condition might have influence on tourist's POI. Third, according to the results, Flickr is not sufficient to conduct sentiment analysis considering positive and negative connotation of tags and descriptions of photos. Textbased social media platforms such as Twitter and TripAdvisor might be useful to conduct such analysis since these platforms are used to voice opinion about places and artefacts. Lastly, official tourism statistics about post-Covid tourism recovery needs to be revised to give further suggestion for policy makers.

#### ACKNOWLEDGEMENTS

Travel expenses to attend the conference were partially covered by the Travel Award sponsored by the open access journal ISPRS International Journal of Geo-Information published by MDPI, and this research was funded by The Republic of Turkiye Ministry of National Education. Authors want to thank Prof. Dr. Bauke de Vries for his suggestions on the study design.

#### REFERENCES

Bihari, A., & Pandia, M. K. (2015). Eigenvector centrality and its application in research professionals' relationship network. 2015 Int. Conf. on Futuristic Trends on Comput. Anal. and Knowl. Management (ABLAZE), 510–514.

CBS. (2021)., The Netherlands in numbers 2021. https://longreads.cbs.nl/the-netherlands-in-numbers-2021/how-many-tourists-stay-overnight-in-the-netherlands

Centre, U. W. H. (2020). Seventeenth-Century Canal Ring Area of Amsterdam inside the Singelgracht. UNESCO World Heritage Centre. https://whc.unesco.org/en/list/1349/

Dane, G., Borgers, A., Kaya, D. I., & Feng, T. (2020). Visitor Flows at a Large-Scale Cultural Event: GPS Tracking at Dutch Design Week. *ISPRS Int. J. of Geo-Information*, 9(11), 661.

Devkota, B., Miyazaki, H., & Pahari, N. (2019). Utilizing User Generated Contents to describe Tourism Areas of Interest. 2019 *First Int. Conf. on Smart Technology & Urban Dev.*, 1–6.

ECIR. (2014). https://ecir2014.org/

ERIH. (2022). https://www.erih.net/i-want-to-go-there/site/heineken-experience-heineken-brewery

García-Hernández, M., de la Calle-Vaquero, M., & Yubero, C. (2017). Cultural Heritage and Urban Tourism: Historic City Centres under Pressure. *Sustainability*, 9(8), 1346.

Ginzarly, M., Pereira Roders, A., & Teller, J. (2018). Mapping historic urban landscape values through social media. *J. of Cult. Herit.*, 36 (2019), pp. 1-11

Girardin, F., Vaccari, A., Gerber, A., & Biderman, A. (2009). Quantifying urban attractiveness from the distribution and density of digital footprints. *Int. J. of Spat. Data Infrastructure*. Res., 4, 26.

Grimaldi, D., Collins, C., & Acosta, S. G. (2021). Dynamic Restaurants Quality Mapping Using Online User Reviews. *Smart Cities*, 4(3), 1104–1112.

Grinberger, A. Y., Shoval, N., & McKercher, B. (2014). Typologies of tourists' time-space consumption: A new approach using GPS data and GIS tools. *Tour. Geogr.*, 16(1), 105–123.

Hauthal, E. & Burghardt, D. (2016). Using VGI for analyzing activities and emotions of locals and tourists. In: Proceedings of AGILE 2016, Helsinki

Het Schip. (2010, March 21). Museum Het Schip. https://www.hetschip.nl/en/about-the-museum/locations/het-schip

Hospers, G.-J. (2019). Overtourism in European Cities: From Challenges to Coping Strategies. CESifo Forum, 20(03), 20–24.

Hu, F., Li, Z., Yang, C., & Jiang, Y. (2019). A graph-based approach to detecting tourist movement patterns using social media data. *Cartogr. and Geogr. Inf. Science*, 46(4), 368–382.

Karayazi, S. S., Dane, G., & Vries, B. de. (2021). Utilizing Urban Geospatial Data to Understand Heritage Attractiveness in Amsterdam. *ISPRS Int. J. of Geo-Information*, 10(4), 198.

Kernel Density. (2022). https://pro.arcgis.com/en/proapp/latest/tool-reference/spatial-analyst/how-kernel-densityworks.htm

Koutras, A., Nikas, I. A., & Panagopoulos, A. (2019). Towards Developing Smart Cities: Evidence from GIS Analysis on Tourists' Behavior Using Social Network Data in the City of Athens. In V. Katsoni & M. Segarra-Oña (Eds.), *Smart Tour. as a Driv. for Culture and Sustain.* (pp. 407–418). Springer Int. Publishing.

Liu, B., Huang, S., & Fu, H. (2017). An application of network analysis on tourist attractions: The case of Xinjiang, China. *Tour. Manag.*, 58, 132–141.

Ministerie van Onderwijs, C. en W. (2019, January 18). https://www.cultureelerfgoed.nl/onderwerpen/rijksmonumenten register/rijksmonumenten-kaartlagen

Quantified Self. (2022). https://www.quantifiedself.nl/

Saxena, A., & Iyengar, S. (2020). Centrality Measures in Complex Networks: A Survey. ArXiv:2011.07190.

Shih, H.-Y. (2006). Network characteristics of drive tourism destinations: An application of network analysis in tourism. *Tour. Manag.*, 27(5), 1029–1039.

Song, H., & Liu, H. (2017). Predicting Tourist Demand Using Big Data. In Z. Xiang & D. R. Fesenmaier (Eds.), *Anal. in Smart Tour. Des.* (pp. 13–29). Springer Int. Publishing.

Trinh, T. T., & Ryan, C. (2017). Visitors to Heritage Sites: Motives and Involvement—A Model and Textual Analysis. *J. of Travel Res.*, 56(1), 67–80.

TripAdvisor. (2022). https://www.tripadvisor.com/Attractionsg188590-Activities-a\_allAttractions.true Amsterdam\_North\_Holland\_Province.html

UNWTO. (2018). Tourism and Culture Synergies. https://doi.org/10.18111/9789284418978

van der Zee, E., Bertocchi, D., & Vanneste, D. (2018). Distribution of tourists within urban heritage destinations: A hot spot/cold spot analysis of TripAdvisor data as support for destination management. *Curr. Issues in Tour.*, 1–22.

Zygomatic. (2016). https://www.wordclouds.com/