

# Analysing tourism movements along the Danube river based on geotagged Flickr photography

Gede, M<sup>a, \*</sup>, Kádár, B<sup>b</sup>

<sup>a</sup> ELTE Eötvös Loránd University, Budapest, Hungary, [saman@map.elte.hu](mailto:saman@map.elte.hu)

<sup>b</sup> Budapest University of Technology and Economics, Hungary, [kadarb@urb.elte.hu](mailto:kadarb@urb.elte.hu)

\* Corresponding author

**Abstract:** The Danube, connecting ten countries with large cities, national parks and other smaller destinations is a branded as integrated destination system of Europe. This paper examines the behaviour of visitors along the river by analysing the spatio-temporal distribution of geotagged photos shared on Flickr. The authors downloaded the metadata of more than 2 million geotagged photos along the river, and created a weighed movement graph based on the short time-interval user movements. The network analysis of the graph revealed the central integrating effect of the large capital cities on the entire system. Modularity analysis indicated 18 regional tourism clusters, which are connected into three major separated destination systems which have little relation between them and the isolated clusters of Lower Danube.

This paper discusses the method of data aggregation as well as the tools used and created for visualising and analysing the data.

**Keywords:** Flickr, geotagged photographs, network analysis, movement analysis

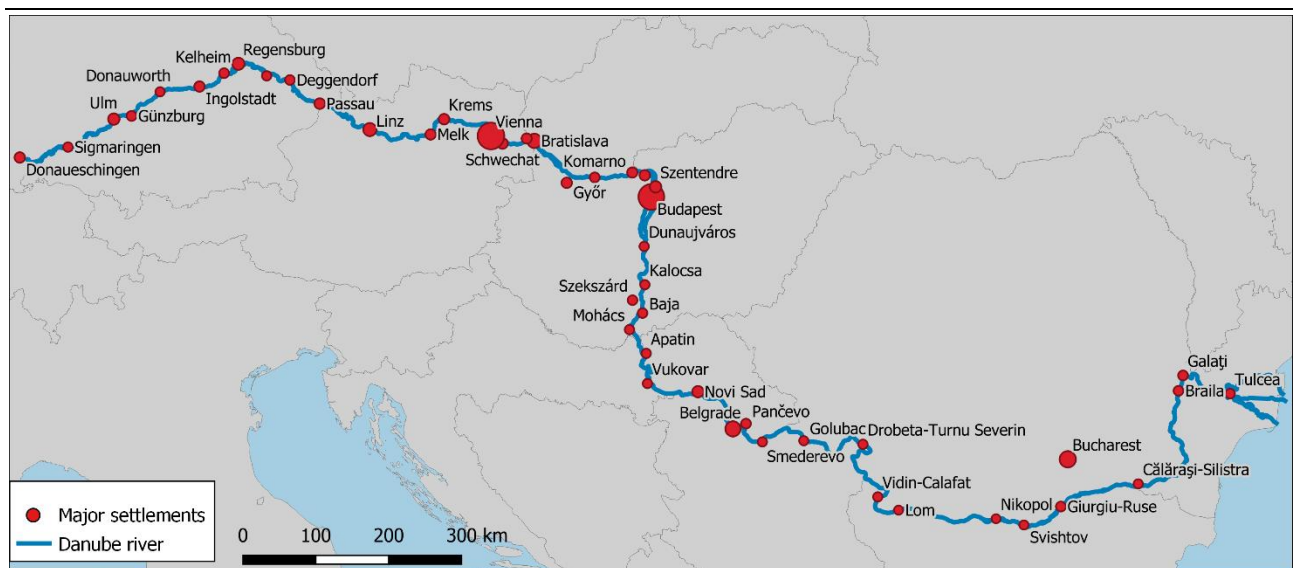


Figure 1. Overview map of the Danube river. Settlement symbol sizes are scaled in respect to number of photographs taken there

## 1. Introduction

The Danube, second longest river of Europe, is an outstanding tourist destination. Along its banks there are four national capitals (Vienna, Bratislava, Budapest and Belgrade), plenty of other historical towns, sixteen protected areas including national parks, nature parks, biosphere reserves (Danubeparks 2019), forming a diverse system of tourist destinations. This paper examines whether these locations form a network, attracting tourists to visit not only one sight at a time but link several attractions on a trip.

These multiple destinations may be in the same region – for example Visegrád, Esztergom, Vác and Szentendre with Budapest, but there are interregional networks as well, connected by various means of transport, including the EuroVelo 6 bicycle route (Meschik 2012) and river cruise ships (Dragin, Jovičić, Lukić 2010).

Tourism is traditionally measured from the supply side – mostly by the number of registered guest nights. Such data is not very useful if more than one single attractions can be visited in one day. User generated content, and social media data and particularly publicly shared photographs became a valuable source of information about various trips. Until recently Flickr.com was the most widely used

web service for photos – unfortunately recently after a change in ownership, this site decreased dramatically the quota of free users (Gartenberg 2018). According to Lo et al. (2011) a considerable part of tourists use this service to share their photographs.

Most tourist photographs nowadays are taken by smartphones or tablets, and usually supplemented by a geotag – a location acquired by the GPS/GNSS unit of the device. This information together with the date/time of the photo makes publicly shared photos a great source of user movement data.

Flickr.com provides an API (Application Programming Interface) that lets users to download metadata of photographs and to analyse it (Popescu, Grefenstette 2009, Gede 2012). Most studies used this information on a local (Kádár, Gede 2013, de Souza et al. 2018) or regional level (Girardin et al. 2008).

The present study aims to retrieve user movement information from Flickr.com on an interregional level, the surroundings of the Danube river.

## 2. Downloading the data

The authors had developed a web based download tool for a previous project (Kádár, Gede 2013), that downloads the metadata of all geotagged public Flickr photos in a given geographic quadrangle and stores it in a relational database. Recently this tool was also implemented as a QGIS plugin (Gede 2018).

This solution, however, is not applicable for the whole Danube as the geographic bounding box of the river covers half Europe (see Figure 1), which would result in a never-ending download process. Instead, the administrative boundaries of all settlements adjacent to the river were extracted from OpenStreetMap data using the Overpass Turbo API (OSM Wiki contr. 2019). The bounding boxes of these polygons were still not optimal for starting the bulk download process, as they had plenty of overlaps. Therefore, they were transformed into a set of adjacent and non-overlapping rectangles (Figure 2) using the following algorithm:

```

bb[] is the array of all bounding boxes
sort bb from west to east
iterate i from 0 to length(bb)-1:
  iterate j from i+1 to length(bb):
    if bb[i] contains bb[j], then remove bb[j]
    else if bb[j] contains bb[i], then:
      remove bb[i]
    end the inner loop
  else if bb[i] overlaps with bb[j]:
    create the overlapping parts of bb[j] as rectangles
    insert them before bb[j]
    remove bb[j]

```

The bulk downloading were performed on all members of the resulting set of bounding rectangles.

The downloaded data contains the metadata (location as lat/lon coordinates, date/time, owner ID, title, tags, photo url) of more than 2 million geotagged photographs along the Danube.

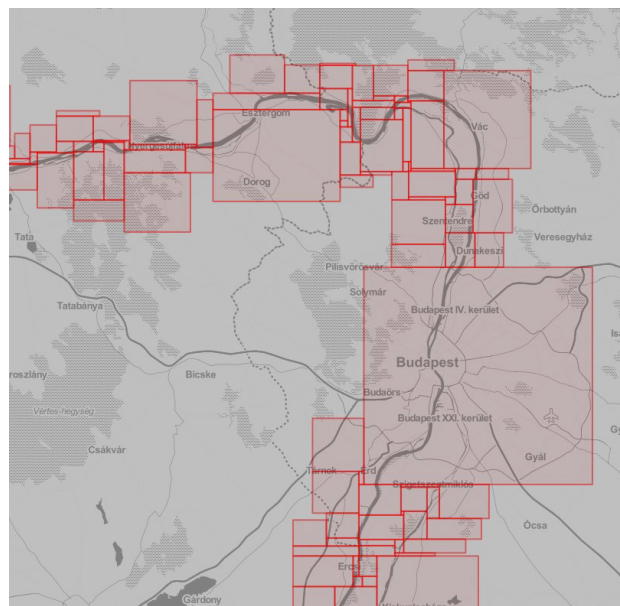


Figure 2. Adjacent, non-overlapping bounding rectangles

This data was completed with the distance from the river streamline, the official river kilometre segment number of the closest point along the river (calculated from the mouth of the Danube Delta), on which riverbank the photo is (left or right), and the settlement the given location belongs to.

## 3. Visualizing the data

An interactive web map application (entitled “Danube Explorer”, available at <http://mercator.elte.hu/~saman/de>) was developed as a tool for further examination of the data. The map indicates the number of photos on each bank by 10km or 1km river segments (depending on zoom level) using variable width colour stripes, red and green for the left and right bank, respectively. Clicking on a segment also reveals the top ten tags attached to photos there (Figure 3).

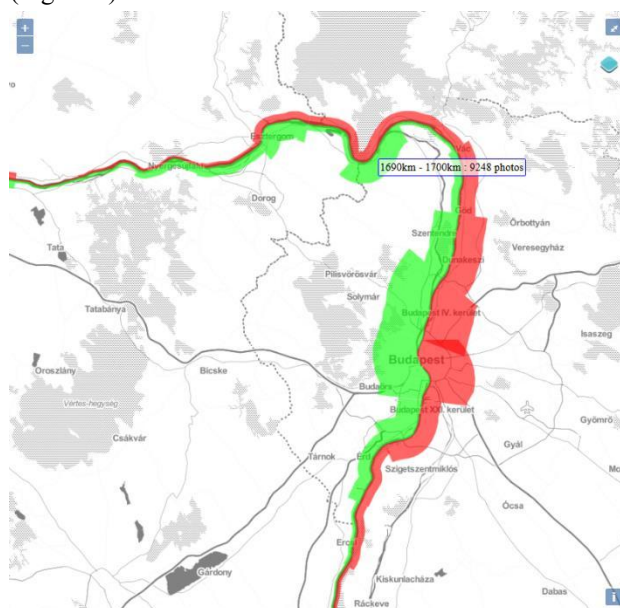


Figure 3. The “Danube Explorer” web map tool.

As a first overview of the photo distribution, a linear diagram of the number of photographs by river kilometres were created (Figure 4). The peaks on this diagram align

with the positions of the major cities and other touristic destinations along the river.

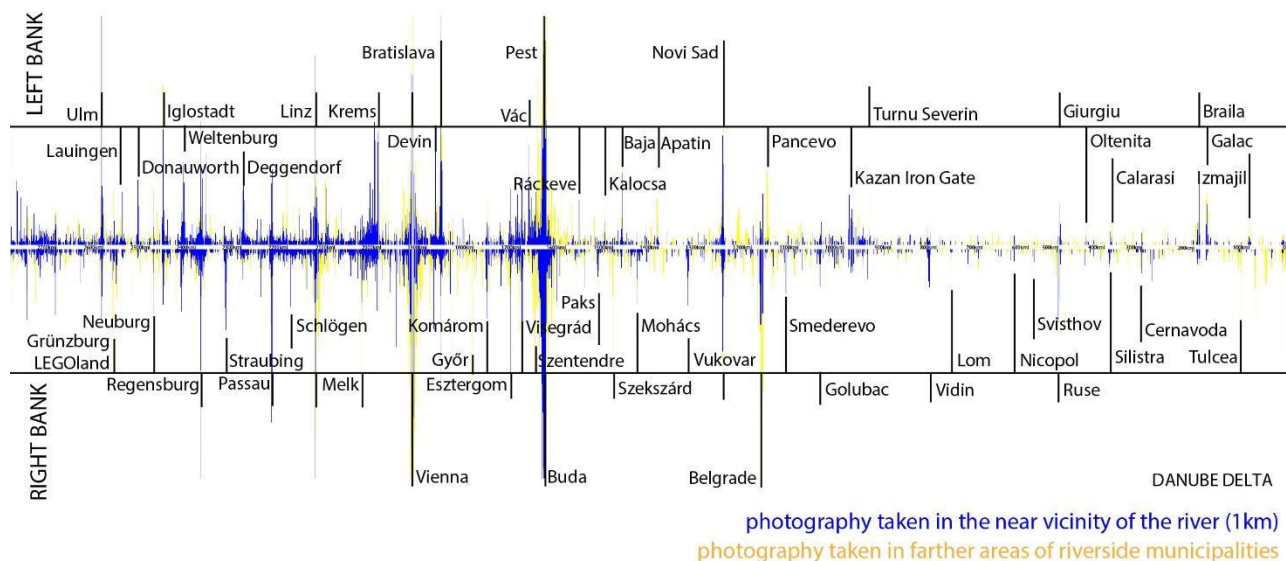


Figure 4. Flickr photo distribution along the Danube, measured in the administrative areas of municipalities adjacent to the river, separated to right and left bank areas. From (Kádár, Gede 2018)

#### 4. Extracting multi-destination visits

As a next step, we tried to reconstruct multiple destination visits along the river. For this, the river was cut into 10km long segments. Some larger cities and other destinations stretched beyond segments limits; these segments were united as the goal was to detect movements between cities not within them.

Photographs were grouped and aggregated based on user ID, the river segment and the date and time (with one-hour granularity), forming a time series of data for each user. Then a “travel graph” was created where nodes are river segments and edges between them indicate that a user made photographs in each node within two days. The graph was elaborated in Gephi, an open source graph analyser software (Bastian, Heymann, and Jacomy 2009). This program can display graphs using geographic coordinates of nodes (“Geo Layout”), therefore these figures can also be superimposed on the map of the area (Figures 5 and 6).

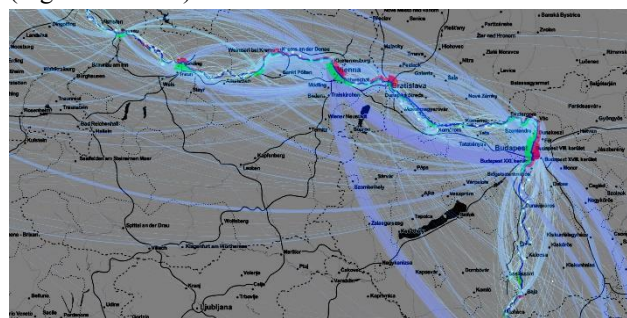


Figure 5. Network graph of destinations superposed on Danube Explorer map. Line width is proportional with the number of users moving between the places. (Detail of map, from Kádár, Gede 2018)

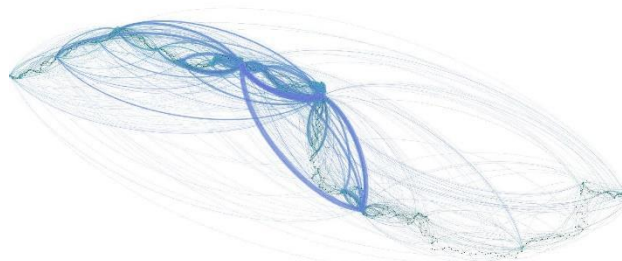


Figure 6. Network graph of destinations. Line width is proportional with the number of users moving between the places. (Full graph, from Kádár, Gede 2018)

#### 5. Cluster analysis of the travel graph

The next step was to create clusters of destinations based on the travel graph. As the connections between the capital cities are larger part not related to the Danube, these segments were excluded from the network. The destination clusters were defined by the modularity analysis of the graph. The weights of edges were set to the number of connections between nodes divided by the geographic distance of them nodes in order to have a weighted network representing travels in a regional scale. The “Modularity” function of Gephi with resolution set to 0.9 was used to find 18 strongly connected clusters. This algorithm uses the Louvain method (Blondel, Guillaume, Lambiotte, & Lefebvre, 2008) which optimizes the partition of the graph to get the maximum possible modularity value (a scalar between -1 and 1 measuring the density of links inside communities as compared to links between communities). The modularity index is  $Q = 0.866$ , which shows a set of well-defined clusters ( $Q=0$  means no separation,  $Q=1$  total separation).



The 18 clusters were united in a new, simplified graph, where these clusters became the nodes and the weight of the edges between two nodes is the sum of the weights of all edges in the original graph between the two given clusters.

## 6. Results

The resulting maps and diagrams show statistically relevant data on the number of visitors to the different sections of the Danube in high details. The “Danube Explorer” map indicates where are the main places of interest along the Danube. Zooming in the accuracy of one kilometre and the division between left and right banks show the exact position of attractions, counting the

number of photographs taken there. This way different sections of the Danube can be compared.

The graph of the destination clusters gives a good overview of the interregional system of tourism along the Danube. The 18 nodes form 3 separated communities, these clusters are to be considered interconnected tourism destination systems, and their boundaries are much defined by national borders. The weight of connections in the Romanian section are definitely lower than in the Upper-Danube regions, nodes here are completely cut-off from the system. The destination system visualized in Figure 7 is the condensed synthesis of all visitor flows measured at a resolution of 10 kms all along the Danube.

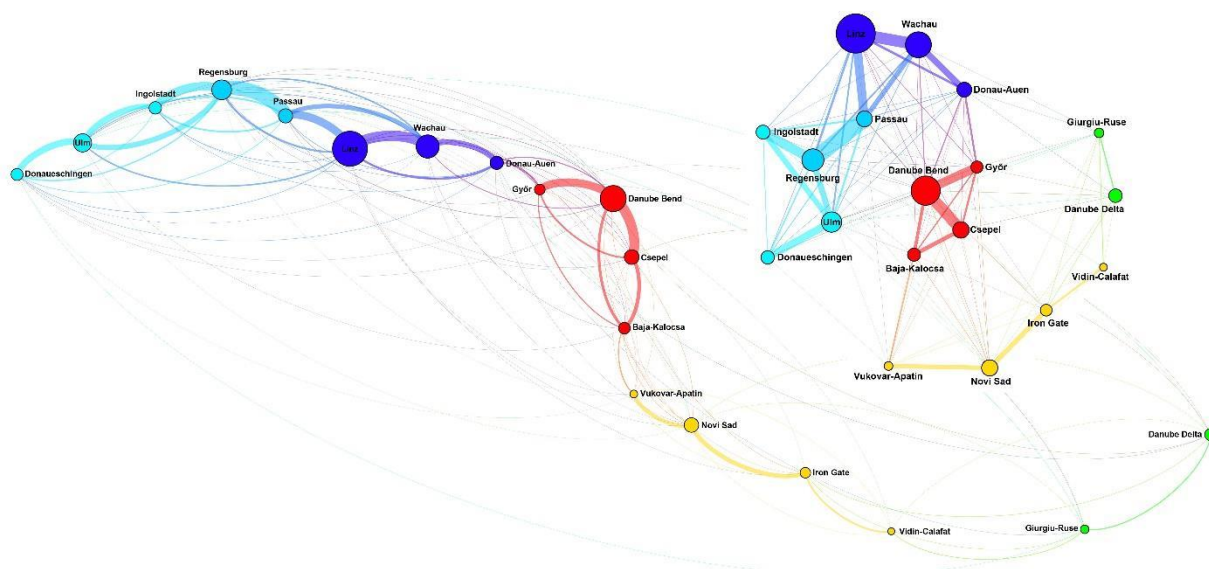


Figure 7. Visitor flows along the Danube between the 18 identified tourism clusters without the five capital cities in a Geo Layout network visualization and Force Atlas 2 layout in Gephi (overlaid graph at top-right). Colours represent the identified cluster groups between tourism clusters.

## 7. References

- Bastian, M., Heymann, S., & Jacomy, M. (2009). Gephi: An Open Source Software for Exploring and Manipulating Networks. *Third International AAAI Conference on Weblogs and Social Media*, 361–362. <https://doi.org/10.1136/qshc.2004.010033>
- Blondel, V. D., Guillaume, J.-L., Lambiotte, R., & Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, 2008(10), 1–12.
- Danubeparks (2019). Danubeparks - Network of Protected Areas. <http://www.danubeparks.org/>
- Dragin, A., Jovičić, D., Lukić, T. (2010). Cruising along the River Danube – Contemporary Tourism Trend in Serbia.” *Geographica Pannonica* 14(3):98–108.
- Gartenberg, C (2018). Flickr will end 1TB of free storage and limit free users to 1,000 photos. *The Verge*, Nov 1, 2018. <https://www.theverge.com/2018/11/1/18051950>
- Gede, M. (2012). Visualization Methods of Spatial Distribution of Geotagged Photography. *Data is Beautiful Conference*. Budapest: Kitchen Budapest. Pp. 52–56.
- Gede, M. (2018). Flickr Metadata Downloader. QGIS Python Plugins Repository. <https://plugins.qgis.org/plugins/flickrddl/>
- Girardin, F., Dal Fiore, F., Blat, J., Ratti, C. (2008). Understanding of Tourist Dynamics from Explicitly Disclosed Location Information. *Journal of Location Based Services* 2(1).
- Kádár, B., Gede, M. (2013). Where Do Tourists Go? Visualizing and Analysing the Spatial Distribution of Geotagged Photography. *Cartographica: The International Journal for Geographic Information and Geovisualization*, 48(2), 78–88.
- Kádár, B., Gede, M. (2018). Mapping the Attractivity of Tourist Sites All Along the Danube Using Geotagged Images from Flickr.com. *5th International Academic Conference on Places and Technologies*, Belgrade, 2018, 766–776.
- Lo, I. S., McKercher, B., Lo, A., Cheung C., Law, R. (2011). Tourism and Online Photography. *Tourism Management* 32(4):725–31.

- Meschik, M. (2012). Sustainable Cycle Tourism along the Danube Cycle Route in Austria. *Tourism Planning and Development* 9(1):41–56.
- OSM Wiki contr. (2019). Overpass API. OpenStreetMap Wiki.  
[https://wiki.openstreetmap.org/wiki/Overpass\\_API](https://wiki.openstreetmap.org/wiki/Overpass_API)
- Popescu, A., Grefenstette, G. (2009). Deducing Trip Related Information from Flickr. *Proceedings of the 18th international conference on World Wide Web*, pp. 1183–84.
- de Souza, T. G., Elias, E. N. N, Fernandes, V. O. (2018) Análise da distribuição espacial a partir de geotagging de fotografias disponíveis na rede social Flickr voltadas ao turismo em Salvador-BA. VII Simpósio Brasileiro de Ciências Geodésicas e Tecnologias da Geoinformação Recife - PE, 08-09 de Nov de 2018.