

UC Berkeley

International Conference on GIScience Short Paper Proceedings

Title

Comparing digital traces of modern travellers to journeys of two 18th-19th century British poets

Permalink

<https://escholarship.org/uc/item/52g534d3>

Journal

International Conference on GIScience Short Paper Proceedings, 1(1)

Authors

Chesnokova, Olga

Gregory, Ian N.

Purves, Ross S.

Publication Date

2016

DOI

10.21433/B31152g534d3

Peer reviewed

Comparing digital traces of modern travellers to journeys of two 18th-19th century British poets

O. Chesnokova¹, I. N. Gregory², R. S. Purves¹

¹Department of Geography, University of Zurich, Winterthurerstrasse 190, 8057 Zürich, Switzerland
Email: {olga.chesnokova; ross.purves}@geo.uzh.ch

²Department of History, Lancaster University, Lancaster LA1 4YD, UK
Email: i.gregory@lancaster.ac.uk

Abstract

The growth of interest in georeferenced media brings with it a wealth of possibilities for exploring digital media, and importantly, for advancing research concerned with domain driven research questions. In this paper, we compare previous work, which extracted and analysed the spatial traces of two 18th-19th century poets in the English Lake District with modern data digital traces in the form of Flickr images. We explore the semantics of the modern day data through use of Latent Dirichlet Allocation, and analyse the extent to which modern day tourism mimics (or indeed follows) the foundations laid by Samuel Coleridge and Thomas Gray. Our results show that tourists, just like the poets, describe mountain landscapes from below and visit popular locations commonly perceived as beautiful.

1. Introduction

The attractiveness of the English Lake District as a destination, both historically and in the present, is reflected by a wealth of poetry, travel diaries, guidebooks, and other forms of contemporary and historical written texts. Cooper and Gregory (2011) recognised the richness of such sources and performed an analysis of texts by two 18th-19th century British poets, Samuel Coleridge and Thomas Gray. In their analysis they explored not only how and what was described, but also the locations visited by the two writers.

Modern day visitors to the Lake District also leave traces, not always in the elegant prose of their forebears, but often through photographs and associated descriptions shared on image hosting web sites such as Flickr or Panoramio, often linked to specific locations through coordinates (c.f. Girardin et al. 2008). These images allow us to explore patterns of landscape description and semantics, and in this paper we explore two questions, and link these to Gray and Coleridge's historical exploration of the Lake District:

1: Do images and metadata allow similar semantic and geometric analysis to that previously carried out by Cooper and Gregory (2011)?

2: Do the spatial traces of modern travellers resemble those of Gray or Coleridge?

2. Methods

2.1 Data preprocessing

Using the Flickr API¹, we retrieved 135649 photographs and their metadata (user, coordinates, time taken and tags) taken by 6855 users in the Lake District. We removed photographs uploaded by single-posters and the most prolific contributors as well as tags with a coefficient of variation greater than 500% (Purves et al. 2011). Furthermore, we also removed tags containing toponyms or camera parameters.

¹ <https://www.flickr.com/services/api/>

2.2 Spatial and semantic analysis

The connection between highly-photographed places (the number of unique users taking a picture at each location) and the attractiveness of places is used in several studies (e.g., Casalegno et al. 2013; Girardin et al. 2008). One way to explore the reasons for this attractiveness is to analyse the underlying semantics captured by tags. We performed an unsupervised classification of tags using Latent Dirichlet Allocation (LDA) (Blei, Ng and Jordan 2003) where a user-defined number of classes from the text is extracted. Each Flickr tag is given a probability of belonging to a class. In this way, the similarities of regions and their typical characteristics can be described.

Several authors have proposed stratifying photographs into those taken by tourists or locals, often using the length of time a user is found in a region. If this time is short, the user is classified as a tourist (Girardin et al. 2008; Straumann et al. 2014). Here we use 30 days to separate tourists from locals.

3. Results

In Figure 1, we compare the distribution of photographs taken by all users (a), tourists (b) and the places mentioned in Gray (c) and Coleridge's accounts (d). Pictures taken by tourists and locals are concentrated around Keswick, Grasmere and Windermere, whereas the regions to the north and east of the Lake District are the least photographed.

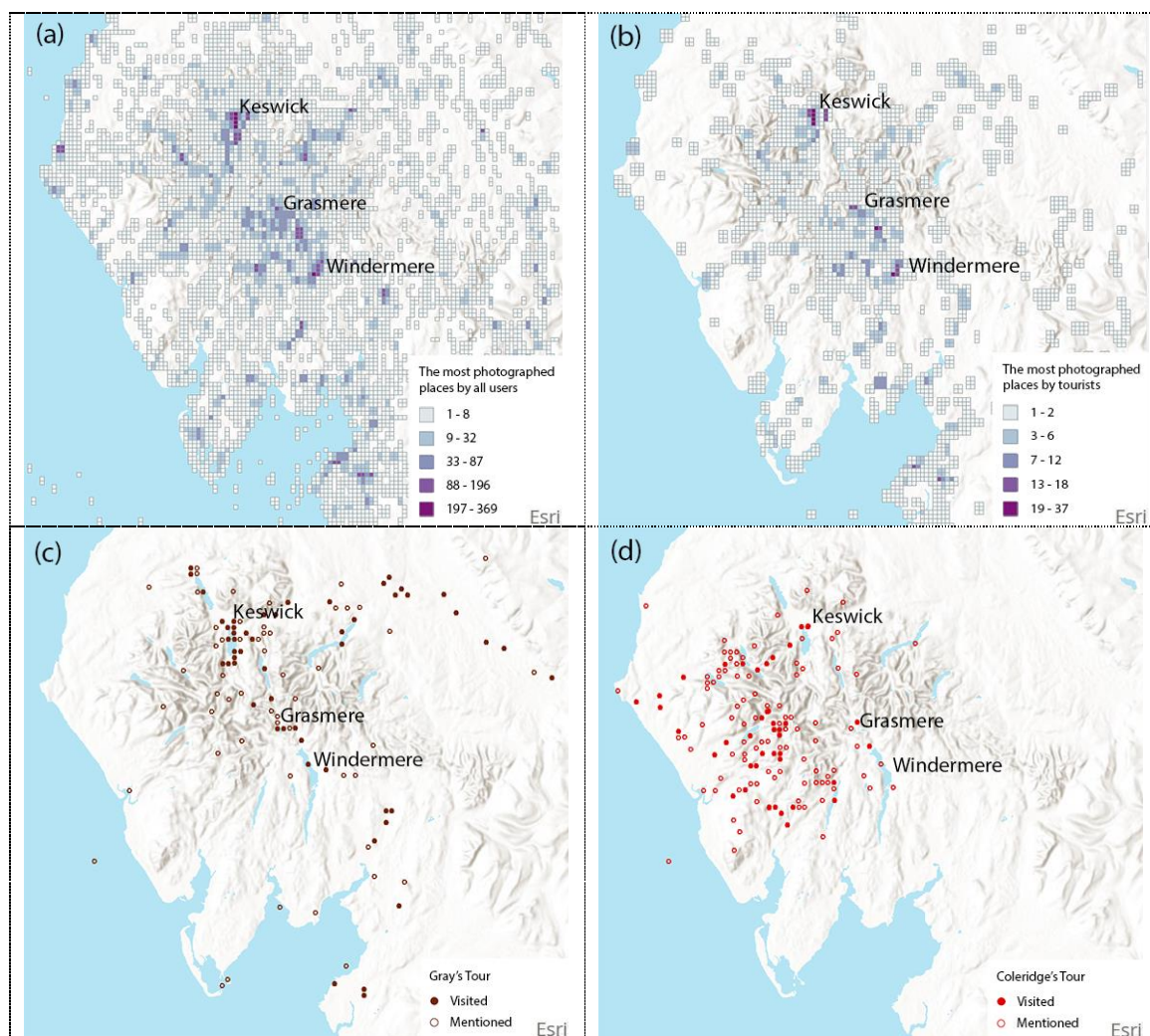


Figure 1. A comparison of the most photographed places by all users and by tourists and the spatial patterns of Gray's and Coleridge's travels.

58% of the photos of today's tourists were taken below 100 m and 92% were taken below 300 m. To draw a comparison with the journeys of Gray and Coleridge, we analysed the altitude associated with descriptions of upland landscapes. To do so we extracted images whose tags described either toponyms found above 600 m or terms such as 'summit,' 'fell,' and 'mountain'. We found, analogously to Cooper and Gregory (2011), that modern day tourists describe the uplands from below, with 21% of upland related descriptions captured below 100 m and 68% below 300 m.

Gray has been described as a "Picturesque" traveller, visiting popular locations commonly perceived as beautiful, while Coleridge was characterised as an "environmental insider and post-Picturesque traveller" (Cooper and Gregory 2011, p. 94) who thus took less well travelled ways. By analysing the number of unique users found within 2 km of places visited by Gray and Coleridge we find that this Picturesque view (as propagated by Gray) of the Lake District continues to be preferred by modern tourists (χ^2 test, $p < 0.05$).

Using LDA, we extracted, and labelled, 10 classes of landscapes from Flickr tags (Table 1). We hypothesised that these classes might reveal relationships between landscapes and observers. For instance, two quite different regions to the north and south are linked through the 'animal class', and turn out to be the locations of wildlife parks. A third location in Penrith features in this class, and a park where red squirrels are commonly photographed is the deciding factor for its membership. Similarly, on lake shores photos of seabirds are influential in linking these locations to this class (Figure 2).

Table 1. Ten classes of landscapes by the LDA unsupervised classification.

Indicative names	The most probable words per class (ranked by probability)
"railway"	railway, boats, train, station, viaduct, steam, hotel, harbour, trains, railways
"animals"	nature, birds, wildlife, north, animal, landscapes, old, spring, animals, northwest
"winter"	winter, snow, rocks, coast, stone, stones, settle, ice, circle, midlandhotel
"holidays"	holiday, sea, beach, sun, architecture, sand, classic, sculpture
"view"	clouds, sunset, countryside, mountains, cloud, jetty, view, scenery, valley, tree
"hiking"	mountains, autumn, walking, hiking, colour, rain, camping, farm, leaves, coastto coast
"summer"	boat, tree, summer, wedding, sunrise, family, pub, places, horse
"nature"	mountain, river, bridge, reflection, stream, village, path, swan, evening, duck
?	grass, sheep, light, reflections, panorama, black, wall, night, field, building
?	walk, seaside, bus, dog, stagecoach, cold, music, long, dusk, frost

The class defined by the tags 'railway,' 'train' and 'station' overlaps with the railway. Different seasons are shown in four colours (Figure 2), but we found no major temporal differences.

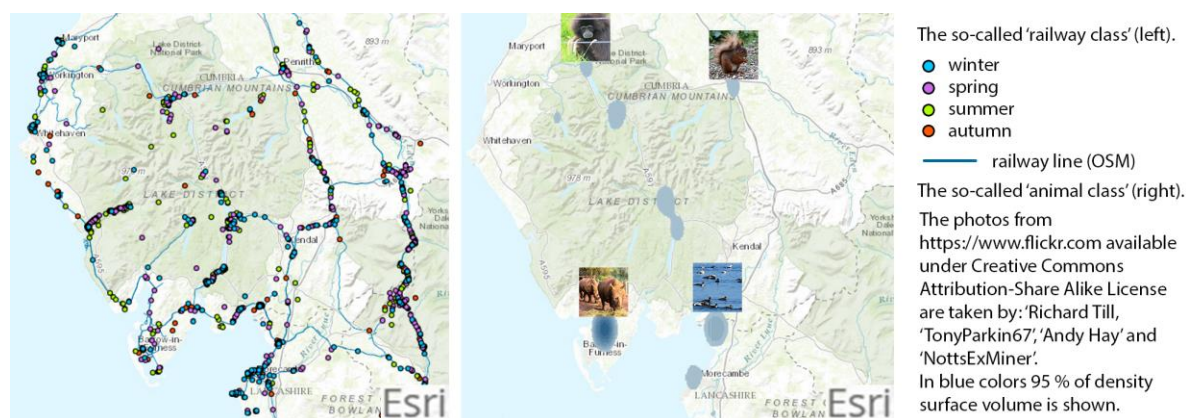


Figure 2. Spatial distribution of results.

4. Concluding discussion

Our motivation was to demonstrate the extent to which it is possible to use Flickr data for the extraction of modern travelling spatial and semantic patterns.

Firstly, our analysis shows that today's tourists have similar travelling patterns to the poets of the 18th-19th century, especially with respect to the Picturesque travelling style of Gray. The villages of Keswick and Grasmere are still considered most attractive and regions in the north remain little visited. Furthermore, the habit of describing the uplands from a low-lying vantage point persists to the present day.

Secondly, we observe that nouns in particular have a high probability in each class (c.f. Purves et al. 2011). This means that we can extract some unexpected patterns as observed with our 'animal class,' but we are not able to define the adjectives that describe the classes (c.f. Cooper and Gregory 2011). An additional difficulty lies in assigning labels to classes, as indicated by question marks in Table 1.

Our results demonstrate that, even with the limited semantic depth available in Flickr images, it is possible to compare not only travel patterns, but also the ways in which landscapes are described, with historical sources. In future work we will extend our approach using richer, natural language, contemporary sources such as travel blogs and guidebooks.

Acknowledgements

Ian Gregory's contribution was funded by the ERC under grant agreement number 283850.

References

- Blei DM, Ng AY and Jordan MI, 2003, Latent Dirichlet Allocation. *The Journal of Machine Learning Research* 3(1):993–1022. doi:10.1162/jmlr.2003.3.4-5.993.
- Casalegno S, Inger R, DeSilvey C and Gaston KJ, 2013, Spatial Covariance between Aesthetic Value & Other Ecosystem Services. *PLOS ONE* 8(6).
<http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0068437>.
- Cooper D and Gregory IN, 2011, Mapping the English Lake District: A Literary GIS. *Transactions of the Institute of British Geographers* 36:89–108. doi:10.1111/j.1475-5661.2010.00405.x.
- Girardin F, Blat J, Calabrese F, Dal FF and Ratti C, 2008, Digital Footprinting: Uncovering Tourists with User-Generated Content. *IEEE Pervasive Computing* 7(4):36–44. doi:10.1109/MPRV.2008.71.
- Purves RS, Edwardes AJ and Wood J, 2011, Describing Place through User Generated Content. *First Monday. Peer-Reviewed Journal on the Internet* 16(9).
<http://firstmonday.org/ojs/index.php/fm/article/view/3710/3035>.
- Straumann RK, Çöltekin A and Andrienko G, 2014, Towards (Re)Constructing Narratives from Georeferenced Photographs through Visual Analytics. *The Cartographic Journal* 51(2):152–165. doi:10.1179/1743277414Y.0000000079.