

# Exploring City Structure from Georeferenced Photos Using Graph Centrality Measures

Katerina Vrotsou<sup>1</sup>, Natalia Andrienko<sup>1</sup>, Gennady Andrienko<sup>1</sup>,  
and Piotr Jankowski<sup>2</sup>

<sup>1</sup> Fraunhofer Institute IAIS, Schloss Birlinghoven, 53757 Sankt Augustin, Germany

<sup>2</sup> San Diego State University, 5500 Campanile Drive, San Diego CA 92182-4493, USA

**Abstract.** We explore the potential of applying graph theory measures of centrality to the network of movements extracted from sequences of georeferenced photo captures in order to identify interesting places and explore city structure. We adopt a systematic procedure composed of a series of stages involving the combination of computational methods and interactive visual analytics techniques. The approach is demonstrated using a collection of Flickr photos from the Seattle metropolitan area.

## 1 Introduction

Photo sharing websites, such as Panoramio ([www.panoramio.com](http://www.panoramio.com)) and Flickr ([www.flickr.com](http://www.flickr.com)), are a popular means of communicating tourist experiences. The posted photographs are freely available, composing a large, rich data source with broad analysis opportunities. Furthermore, they are associated with geographical coordinates and capturing times (geo- and time-referenced), making it possible to regard users' postings as sequences of visited places over time and treat them as trajectories describing their movement. The vast size and multi-faceted nature of such data makes their analysis both challenging and interesting.

We explore the potential of applying graph theory measures of centrality to the network of movements extracted from georeferenced photo sequences in order to identify interesting places and explore city structure. To accomplish this we have defined a procedure composed of a series of stages and involving the combination of computational and interactive visual analytics techniques, including clustering, graph analysis, and aggregation.

The analysis of people's activities and movement using georeferenced photographs has been the subject of previous research [5,4,7,6]. The combination of several algorithmic and visual techniques into a concretely defined procedure is what makes this approach unique. The entire procedure is incorporated into a powerful visual analytics framework which allows a user to interactively investigate the results, alter parameters, re-run the process and make refinements.

## 2 Visual Analytics Approach

The approach we propose involves the retrieval of a graph of movements from a collection of georeferenced photographs captured across a city, and the application of graph theory methods in order to analyse the structural connectivity

of this graph and extract information about the structure of the city itself. The procedure we have defined for accomplishing this is composed of four stages, each of which involves several steps performed systematically:

1. *Pre-processing.* Trajectories are extracted from georeferenced photos.
2. *Aggregation of trajectories.* The extracted trajectories are aggregated into moves between generalized places.
3. *Graph of moves and place centrality.* The graph of aggregated moves is considered and its connectivity is assessed by computing centrality scores.
4. *Analysis of results.* The structural connectivity of the graph is interactively analysed using both visual and algorithmic methods.

## 2.1 Pre-processing

The collection of photographs is first transformed into a set of trajectories. Photographs entered to a photo sharing website by the same person are ordered chronologically forming a sequence of photographed places representing the trajectory of this person in geographical space. Entries that are not properly georeferenced or do not include a time-stamp, and single entry sequences are ignored. Trajectory analysis methods can also be applied for separating trajectories of tourists and locals in order to reveal different characteristics of a city.

## 2.2 Aggregation of Trajectories

The extracted trajectories are aggregated into flows between generalized places using the method presented in [2] which is performed through the following steps:

1. Characteristic points are extracted from the trajectories.
2. Spatially-bounded density based clustering is applied to group the points by their spatial proximity.
3. The centroids of the retrieved clusters are used as seeds for generating Voronoi polygons.
4. The polygons define the set of *places* that the explored area is divided into and reflect the spatial density of the trajectories.
5. The trajectories are aggregated into *moves* between pairs of *places* by defining transitions between them, and counting the number of transitions present.

Clustering parameters, such as the minimum number of elements and the desired radius of spatial clusters, can be interactively adjusted and thus influence the size and number of extracted places.

## 2.3 Graph of Moves and Place Centrality

The aggregation results of the photographers' trajectories can be represented using a directed, weighted graph,  $G = (V, E)$ , where vertex set,  $V$ , is the set of extracted places, and edge set,  $E$ , is the set of directed edges  $(i, j)$  corresponding to moves from place  $i$  to place  $j$ , each one weighted by the number of moves present. The structural importance of the vertices in the graph, and hence of the

extracted places, can be studied by examining their centrality. Several measures of centrality exist for describing different aspects of this structure. We compute and use the following measures (refer to [3] for details):

1. *Degree centrality* of a vertex is defined by the number of its connecting edges. Given our directed, weighted graph we consider four measures of degree centrality: (1) *in-* and (2) *out-degree* are the number of edges connecting into and out of a vertex respectively, and (3) *weighted in-* and (4) *weighted out-degree* are the sum of weights connecting into and out of a vertex.
2. *Closeness centrality* is based on the distance, measured as the size of the shortest path, of a vertex to all other vertices. Closeness centrality of vertex  $v$  is defined as the inverse of the total distance of  $v$  to all other vertices.
3. *Betweenness centrality* measures centrality by considering the number of shortest paths that pass through a vertex,  $v$ , and is defined as the proportion of shortest paths between vertices  $u$  and  $s$  that pass through  $v$ .
4. The *clustering coefficient* represents the likeliness that two neighbours of a vertex  $v$  are connected and measures in a sense the importance of a vertex in its immediate neighbourhood. The clustering coefficient of vertex  $v$  is defined as the fraction of  $v$ 's neighbours that are also neighbours of each other.

## 2.4 Analysis of Results

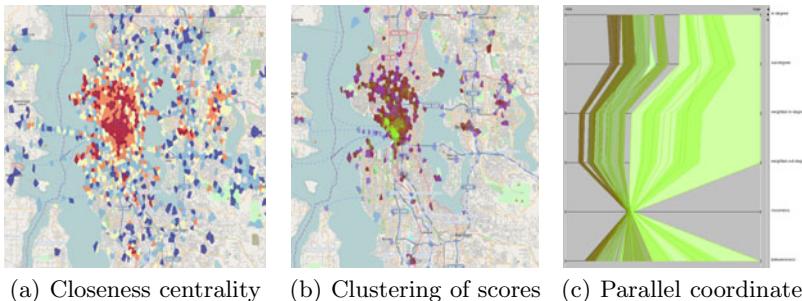
The analysis of the computed centrality is performed through a number of interactive visual analytics tools [1]. Centrality scores are explored, separately or simultaneously, using different visual representations:

- Classified choropleth maps are used to display the scores separately.
- Diagrams are overlaid on a map to make relative comparisons between the centrality scores of places.
- Parallel coordinate plots are used to explore and compare scores of places.
- Clustering of places with respect to their scores is performed to study the spatial distribution of nodes with similar characteristics.
- Dynamic filtering of *places* and *moves* based on centrality is applied to discover interesting places on the map.
- Time graphs and animated maps are used to display time intervals of arbitrary length in order to explore the behaviour of the graph over time.

## 3 Demonstration: Seattle Metropolitan Area Photos

We apply the presented visual analytics procedure to a collection of Flickr photos from the Puget Sound metropolitan area, Washington State, USA. The data were pre-processed and a dataset was retrieved of 577,053 photos captured by 9,324 photographers, between January 1, 2005 and August 31, 2009.

During the aggregation stage a radius of 500 meters was used for clustering the photos by spatial proximity resulting in the extraction of 2,899 places. Places visited less than 5 times were not considered, resulting in a graph consisting of 1,446 vertices (places) and 40,627 edges, each of them weighted by the number



**Fig. 1.** Representation examples from the exploratory analysis of the Seattle metropolitan area. (a) Closeness centrality in a classified choropleth map. (b) Results of clustering places with respect to centrality scores in a choropleth map. (c) Parallel coordinates plot of centrality scores of four clusters coloured by their corresponding cluster colour.

moves between connecting places. Centrality scores were computed for each of the places in the graph.

Several interesting observations were made during the exploratory analysis of this Seattle dataset, these will be demonstrated in their entirety at the conference (a sample is available at <http://geoanalytics.net/and/slides/pkdd11.pdf>). Examples include the interesting picture of the core and periphery of well connected locations revealed by the high closeness values (fig. 1(a)), the identification of the most well connected places through clustering with respect to centrality scores (fig. 1(b)), and the exploration of the centrality score values themselves using different representations and filtering tools (fig. 1(c)).

## References

1. Andrienko, G., Andrienko, N., Bak, P., Keim, D., Kisilevich, S., Wrobel, S.: A conceptual framework and taxonomy of techniques for analyzing movement. *Journal of Visual Languages & Computing* 22(3), 213–232 (2011)
2. Andrienko, N., Andrienko, G.: Spatial generalization and aggregation of massive movement data. *IEEE Transactions on Visualization and Computer Graphics* 17(2), 205–219 (2011)
3. Brandes, U., Erlebach, T. (eds.): *Network Analysis*. LNCS, vol. 3418. Springer, Heidelberg (2005)
4. Crandall, D.J., Backstrom, L., Huttenlocher, D., Kleinberg, J.: Mapping the world's photos. In: 18th International Conference on World Wide Web, pp. 761–770 (2009)
5. Girardin, F., Calabrese, F., Fiore, F.D., Ratti, C., Blat, J.: Digital Footprinting: Uncovering Tourists with User-Generated Content. *IEEE Pervasive Computing* 7(4), 36–43 (2008)
6. Jankowski, P., Andrienko, N., Andrienko, G., Kisilevich, S.: Discovering Landmark Preferences and Movement Patterns from Photo Postings. *Transactions in GIS* 14(6), 833–852 (2010)
7. Kisilevich, S., Krstajic, M., Keim, D., Andrienko, N., Andrienko, G.: Event-Based Analysis of People's Activities and Behavior Using Flickr and Panoramio Geotagged Photo Collections. In: 14th Int'l Conf. Information Visualisation, pp. 289–296 (2010)