Computers, Environment and Urban Systems 53 (2015) 96–109

Contents lists available at ScienceDirect

Computers, Environment and Urban Systems

journal homepage: www.elsevier.com/locate/compenvurbsys

# A multi-scale approach to exploring urban places in geotagged photographs

Rob Feick [a,*], Colin Robertson [b]

[a] School of Planning, University of Waterloo, 200 University Avenue West, Waterloo, ON N2L 3G1, Canada
[b] Department of Geography & Environmental Studies, Wilfrid Laurier University, 75 University Avenue West, Waterloo, ON N2L 3C5, Canada

ARTICLE INFO

*Article history:*
Available online 10 January 2014

*Keywords:*
User generated content (UGC)
Volunteered Geographic Information (VGI)
Geotagged photographs (GTP)
Local log-odds ratio
Multi-resolution
Flickr

ABSTRACT

User-generated content (UGC) that contains spatial references, often referred to by the more bounded concept of Volunteered Geographic Information (VGI), is often touted as a potentially revolutionary data source for geographical research. This paper explores the capacity of one increasingly prevalent source of these data, geographically encoded photographs, to capture spatial expressions of place in an urban environment. Geotagged photographs were obtained from the Flickr API to build a geographic database of photographs for the city of Vancouver, Canada from 2001–2012. These data were aggregated to multiple geographic units represented as hexagonal lattices. Spatial patterns of photo aggregation were examined for tessellations that ranged from 0.25 ha to 1024 ha. Tags associated with each photo were also explored through the notion of 'tag-space' at multiple resolutions, or "scales", of analysis through local log-odds ratios. Results indicate a significant interaction between tag-space semantics and spatial aggregation which suggests that consideration of scale effects should be integral to analysis of this type of tagged VGI for exploring citizens' sensing of urban environments. The results indicate further that we may have to reconsider the interaction between encoded meaning, the methods used for extracting such meaning from tag-space, and exogenous and endogenous spatial scales of spatial UGC.



## 1. Introduction

User-generated content (UGC), information or media that citizens use, create, and share online, has evolved from what many viewed as a curiosity or a passing technology-led fad in the early days of the Web 2.0 era to now being recognized as an important element of many government, business, scientific and social processes (Elwood, Goodchild, & Sui, 2012; Hermida, 2010; McKenzie et al., 2012; Shirkey, 2008; Tapscott & Williams, 2006). Increasingly, the production and communication of information and knowledge are becoming more collaborative and rooted in networked communities as the distinctions between data users and data producers and, by extension, experts and amateurs are blurred. A small, but growing proportion of UGC contains geographic information either in the form of explicit spatial coordinates generated from personal locational devices (e.g. mobile phones, GPS units, etc.) or less explicit references such as names of landmarks, regions or cities.

Some forms of spatial UGC, such as users' mapping of animal sightings, drivers' reports of potholes in roads, and citizens' comments on land management issues, have direct linkages to specific data products, community interests or "citizen science" initiatives (Connors, Lei, & Kelly, 2011; Wiersma, 2010). This spatial subset of UGC is often referred to within the GIS and GIScience literature as Volunteered Geographic Information (VGI) community following Goodchild (2007). The linkages are less evident with other forms of spatial UGC (e.g. photos, videos and Twitter posts with geographic references) that are typically created simply as an outflow of personal interests or web-based communication and may not be not viewed by their authors as geographic data in their own right (Feick & Roche, 2012; Purves, Edwardes, & Wood, 2011). With much of the social web now enabled with location sensors, the social web is fast becoming the social geo-web (Sui & Goodchild, 2011), and geographers and others have expressed great interest in the opportunities and dangers associated with these vast new sources of data (e.g. Wilson, Gosling, & Graham, 2012).

A prime example of the spatialization of social-web interactions is the proliferation of production and access to digital photographs encoded with geographic information and the subsequent exploitation of these repositories for exploratory, national and global-level research (Crandall, Backstrom, Huttenlocher, & Kleinberg, 2009; Zhang, Korayem, You, & Crandall, 2012). Much of this research has centerd on exploring what types of information can be derived from the growing volumes of geotagged photographs (GTP) that are uploaded and shared on sites such as Panoramio, Instagram and Flickr (Antoniou, Morley, & Haklay, 2010; Kennedy,

* Corresponding author. Tel.: +1 519 888 4567x35493; fax: +1 519 725 2827.
*E-mail address:* rdfeick@uwaterloo.ca (R. Feick).

0198-9715/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.compenvurbsys.2013.11.006

Naaman, Ahern, Nair, & Rattenbury, 2007). GTPs are comprised of a georeferenced image and a set of descriptive keywords or tags that users add to describe the photo, the place or event it is associated with, or its personal meaning. Since users determine what they photograph, what tags they use, and which photos they share, GTPs hold promise as a rich data source for investigating how people perceive and characterize the phenomena they photograph and, potentially, for uncovering hidden geographies and spatial structures in social processes (Ferarri, Rosi, Mamei, & Zambonelli, 2011).

In this paper, we propose a multi-scale method for exploring patterns in the spatial and thematic content of GTPs. Our intent is exploratory in nature and is directed at distilling new insights or hypotheses concerning localized, or scale-dependent, expressions of place that are encoded within individuals' GTPs. We build upon a sizeable body of research centered on GTPs including recent work that investigate: concepts of place and vernacular geographies across space and time in Flickr and Geograph (geograph.org.uk/) photographs (Dykes, Purves, Edwardes, & Wood, 2008; Hollenstein & Purves, 2010; Purves et al., 2011), relationships between physical and cyber spaces (Graham & Zook, 2011), individuals' patterns of movement in urban environments (Andrienko & Andrienko, 2011; Jankowski, Andrienko, Andrienko, & Kisilevich, 2010; Kisilevich, Keim, & Rokach, 2010) and how tagged social media can be mined to discern non-experts' approaches to classifying images (Rorissa, 2010) or social trends (Jin, Gallagher, Cao, Luo, & Han, 2010), among others.

A significant thread in the research on GTPs has focused on developing and evaluating measures of tag similarity in order to characterize massive numbers of GTPs (Crandall & Snavely, 2012; Crandall et al., 2009) or to support online tools such as tag recommendation engines, query tools, and visualizations of tag semantics (Moxley, Kleban, & Manjunath, 2008; Wu, Hua, Yu, Ma, & Li, 2008). Indeed, much of the research to date using GTPs has been computational in nature, rather than analytically focused, although some recent examples show considerable promise (e.g. De Choudhury et al., 2010). We suggest that the potential exists to generate a finer understanding of *local* environments through investigation of the spatial, temporal, and semantic qualities of GTPs. Rattenbury and Naaman (2009), for example, demonstrate an innovative approach for distilling place semantics from GTP collections using tag- and spatial-scanning approaches that identify locations where significant concentrations or "bursts" of place-related tags are found.

We approach this dimension of localized characterization of space and place by applying a metric adapted from classical odds ratios, used in ecology and epidemiological studies to measure exposure effect, to search for "tag-space" neighborhoods within an urban environment. We are interested here in exploring whether an ecological approach can help quantify the intrinsic spatial and thematic properties of GTPs and uncover areas of tag similarity. In contrast to much of the other research of this type, we use a multi-resolution approach to GTP aggregation that, we suggest, facilitates exploration of urban patterns and processes that occur across different geographies or "scales" of analysis (e.g. streetscapes, neighborhoods, regions). We believe that this approach has potential to help us to understand, to some degree, the strength and spatial extents of citizens' perceptions, filtering and cognition of urban processes and forms. In this way, we may be able to situate previous GTP research related to spatializing place via space–time patterns in Flickr photos (e.g. Crandall et al., 2009; Kennedy et al., 2007) within a broader context of small scale patterns associated with GTPs. Additionally, large scale studies such as those tracing individuals' movement through urban space from temporal trajectories of GTPs (e.g. De Choudhury et al., 2010; Jankowski et al., 2010) may fit another class of scale-specific processes associated with certain landscapes.

In the following pages, we demonstrate our approach to GTP analysis using data obtained for Vancouver, Canada from the Flickr API. Prior to discussing the methods used, we first describe the study area selected for analysis and provide some context related to spatial and temporal changes to the study area during the study period. We then describe our data processing methods and the building of a final GTP database in detail. Following a description of the data analysis methodology, we provide a series of results examining trends in tagging of GTP across space and using tessellations of varying resolutions. We conclude the paper with a discussion of our key findings and highlight caveats of the current analysis and areas for further research.

## 2. Methods

In this analysis, we are interested in the joint analysis of space, scale and meaning embedded in GTPs obtained for our study area. Scale is handled in a straightforward manner, by making observations at numerous spatial aggregations and comparing how measures change across scales, as is common in biogeographical and ecological research (e.g. Fortin & Dale, 2005; Turner, Gardner, & O'Neill, 2001). Note that in this context, the term "scale" is used as an indication of the resolution at which space is subdivided such that large scale measurements apply to small geographic areas (e.g. points of interest, streetscapes, etc.), while broad regional trends or processes are referred to as small scale (e.g. city-wide commuting flows). Meaning, in the context of analyzing GTPs, requires measurement of either similarity in tag-space or content-space. Tag-space analysis often encompasses tag co-occurrence measures as previously described and examining how these measures are spatially patterned. Hollenstein and Purves (2010) and Jankowski et al. (2010), for example, mine tag descriptions of Flickr GTPs to explore the spatial extent of city cores based on the spatial concentration of selected place terms and landmark preferences respectively. Content-space analysis involves analysis of the digital content of individual and collections of photographs to summarize their relative similarities along some dimension of comparison. For example, an aspatial content analysis called Flickr Distance represents one way this might be accomplished, through clustering of visual language model summaries of pairs of images with common tags (Wu et al., 2008). However relating GTPs based on geography and image content is problematic because people often photograph objects unrelated to their geographic locations. An alternative GTP repository, such as Geograph, whereby participants are tasked with photographing landscapes for specific grids of a reference map covering all of England, holds promise for supporting geo-content analysis, albeit with a sacrifice in generality and spatial resolution (see Dykes et al., 2008; Purves et al., 2011, for example). The approach taken in this paper to elicit meaning is to examine two properties of GTPs across geographic locations: GTP magnitude and GTP tag similarity. Ultimately we aim to associate measures of these properties with meaningful forms and processes in urban environments.

### 2.1. Data and study area

Data were obtained from the Flickr API for the years 2001–2012 using a radius search at fixed locations in the City of Vancouver. The City of Vancouver is an urban center of 600,000, with approximately 2.1 million in the greater region of Metro Vancouver. The City of Vancouver proper is a growing urban center with heterogeneous cultural and physical landscapes. The physical constraints on the city – an international border to the south, mountains to the north, and ocean to the west – have dictated rapid cycles of urban redevelopment and densification in and around the urban core. The downtown core is representative of the post-industrial city; a result of dis-investment and capital flight in the 1980s, and subsequent real estate investment and redevelopment in the 1990s which continues today (Beasley, 2000; Kear, 2007). Vancouver is an interesting example of urban core redevelopment as it has been extremely successful in attracting both residents and visitors to the downtown area. As much of the influx of capital was from Asia, the social fabric underwent transformation during this period, with massive changes in the ethnic makeup of the population. The transformation of Vancouver throughout the 1990s and 2000s was facilitated by a regional planning strategy centered on developing an urban core around principles of liveability: urban design, walkability and sustainable transportation (Beasley, 2000). As a result of this focused redevelopment strategy coupled with its natural beauty, Vancouver has been rated as one of the most liveable cities in the world many times (Economist Intelligence Unit, 2012). What has resulted is thus a mix of the old and new, in physical and human terms, with many new residents drawn to the city for its liveable characteristics and many areas revitalized as urban parks, and transit-oriented centers of development. This urban fabric therefore provides a rich milieu for exploring the varied ways and spatial patterns in which people photograph their environment. Exploring patterns of variability in locations of photograph frequency and how people characterize (i.e. tag) associated photographs may provide insight into how people perceive places across the city.

Flickr data were obtained via the Flickr API using a python script. After some experimentation with the API, a grid of reference points spaced 5 km apart was generated for the study area (Fig. 1), and a radius search was used at each grid point location to obtain nearby photographs. The API was accessed via the python library flickrapi (available from http://stuvel.eu/flickrapi) which enabled extraction of selected metadata for each Flickr photo including the spatial coordinates, date and time of capture, reported accuracy, user-specified tag sets, and the URL of each GTP. Data were constrained to the City of Vancouver boundaries and the time period of January 1, 2001 to December 31, 2012. Once duplicate photographs were removed, 54,522 photos with unique Flickr photo IDs and valid latitude and longitude values were used for the analysis. A total of 4666 unique Flickr contributors were represented in this dataset.

### 2.2. Spatial data processing

Individuals who create and subsequently share VGI, such as GTPs available on public photo sharing resources like Flickr, do so for many reasons including personal satisfaction, altruism, improved professional standing, and enhancement of skills, among others (see Coleman, Georgiadou, & Labonté, 2009). More importantly in the context of this paper, this generation of content is the product of citizens' efforts to sense, sample and document a variety of events, processes and features characteristic of urban environments. To explore the spatial dimension of these processes of sensing, sampling and documenting, the Flickr data described above were aggregated using a hierarchy of geographic resolutions. Our choice of resolutions was conditioned in part by the urban focus of the analysis and in recognition of the long history of examining urban patterns and processes across geographic units that range from micro or site-specific scales (e.g. traffic accidents, etc.), through meso scale (e.g. street, neighborhood) to regional or macroscale foci (e.g. commuter travel patterns, long range land use management) (Sheppard & McMaster, 2004). Our choice to use a multi-resolution approach to GTP aggregation was also influenced by the uncertainties inherent to GTP locations that relate to: (a) method of coordinate capture (e.g. automated encoding of GPS coordinates in a photograph's EXIF metadata versus manual placing of a photograph's location with reference to a web base map, etc.), (b) differences in users' expertise in geotagging, and (c) the relationship between geotagged location and its focal point

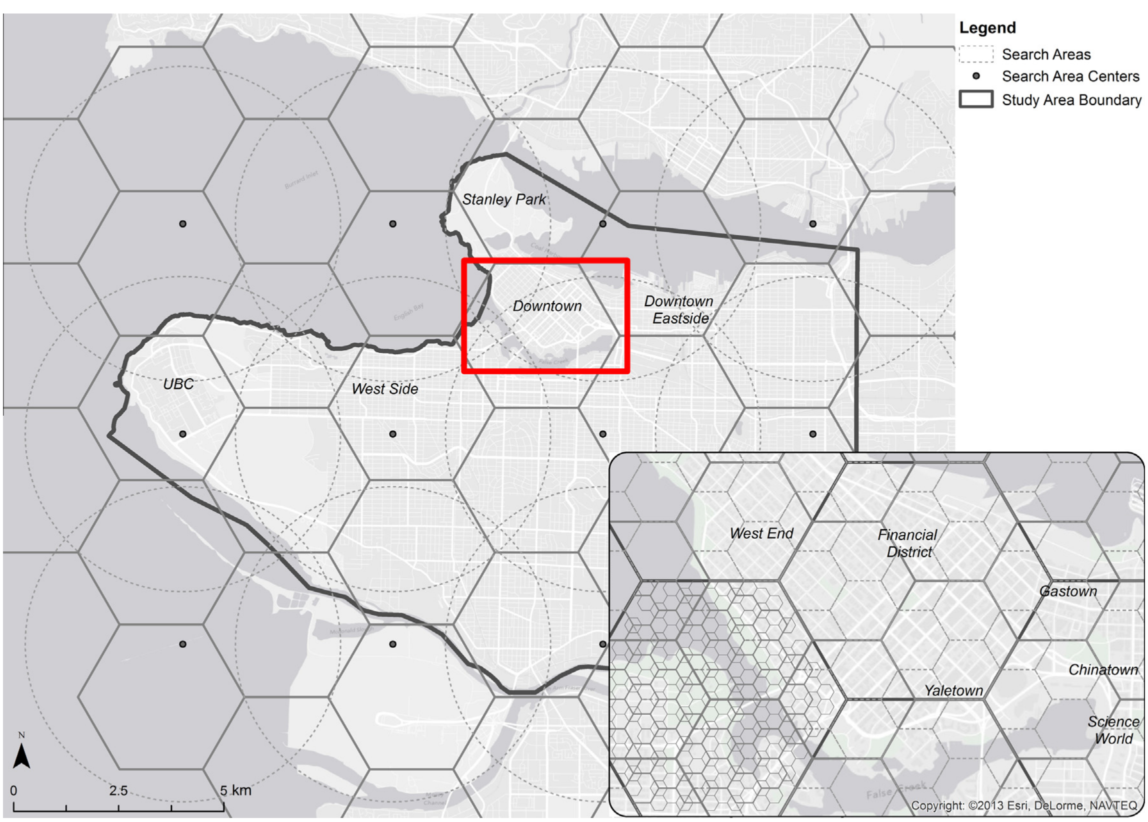


**Fig. 1.** Vancouver study area and nesting of hexagon levels 1024, 256 and 4 ha.

(i.e. do the GTP coordinates represent where the photo was recorded or the location of the photo's subject?) (Hollenstein & Purves, 2010; Jankowski et al., 2010).

In recognition of these considerations, hexagon tessellations of the Vancouver study area were created with cell sizes that range in area from 0.25 ha (.0025 km$^2$) to 1024 ha (10.24 km$^2$) that were seen to approximate local urban places, street scale, neighborhoods and larger districts characteristic of larger urban centers (see Table 1). Hexagon tessellations are typically chosen as sampling units for multiscale spatial analysis (Patil, Myers, Luo, Johnson, & Taillie, 2000; White, Kimerling, & Overton, 1992) due to their topological and geometric properties such as compact shape and simple recursive hierarchical nesting. Further, the neighborhood properties (e.g. 6 equal neighbors for all non-edge hexagons) are conducive to the local neighborhood analysis approach taken in this analysis (Boots & Tiefelsdorf, 2000). Boots and Tiefelsdorf (2000) specifically suggest that hexagon tessellations, on average, are more similar topologically to real-world irregular partitions (e.g. regions, neighborhoods), and as such should be used to investigate spatial patterns in empirical maps. Hexagons were generated as ESRI shapefiles using Rempel, Kaukinen, and Carr (2012) Patch Analyst add-in for ArcGIS software such that higher level hexagons contained three complete hexagons from the next lowest level along with one-third of three neighboring hexagons. Fig. 1 provides a partial illustration of this nesting across the study area with respect to hexagons at the 1024, 256 and 4 ha sizes.

The seven hexagon shapefiles generated by Patch Analyst and a comma-separated file (csv) that contained the Flickr GTP data were imported as separate spatial tables within a PostgreSQL database that was equipped with Refractions Research's PostGIS extension. PostGIS provides Open Geospatial Consortium (OGC) compliant spatial database functionality to the PostgreSQL database and enables storage and management of spatial geometries as well as permitting geographic queries and transformations to be conducted through SQL statements (see http://postgis.net). The performance, open source nature of PostgreSQL and PostGIS, coupled with support of several programming languages for task automation and extending functionality have contributed to increasing use of this combination as a spatial database for a variety of geospatial analysis and web-mapping applications (Chen & Xie, 2008).

The general procedure used to process the Flickr GTP data and permit multi-scale exploration of spatial and tag-space patterns is summarized first and subsequently elaborated upon. In brief, the following procedure was used for each tessellation (i.e. $h_{(25)}$ to $h_{(1024)}$) listed in Table 1:

(a) Spatially join the GTP points to hexagon layer $h_{(i)}$ and record the hexagon ID each photo was found within.
(b) For each hexagon in $h_{(i)}$, count the number of GTPs, the number of unique photographer IDs, identify the 10 most frequently occurring tags and count their frequencies.
(c) For each hexagon in $h_{(i)}$, count the frequency of tag $t$ in the first-order neighboring hexagons.
(d) Calculate the tag-similarity ratio of each hexagon relative to its neighbors as described in equation 1 below.

**Table 1**
Hexagon sampling tessellations.

| Area (ha) | Approx. width (m) | Approx. diagonal length (m) | Number of hexagons |
|---|---|---|---|
| .25 | 54 | 62 | 120,896 |
| 1 | 107 | 124 | 30,456 |
| 4 | 215 | 248 | 7739 |
| 16 | 430 | 496 | 1980 |
| 64 | 860 | 993 | 518 |
| 256 | 1719 | 1995 | 150 |
| 1024 | 3438 | 3970 | 44 |

PL/pgSQL scripting was used to automate this procedure within the PostGIS database. For each hexagon tessellation, several steps were taken to process the Flickr data. For mapping convenience, all Flickr data were initially stored as fields in the seven spatial hexagon tables. For example, all of the unique tags found on any of the Flickr GTPs that fell within a given hexagon were assembled using a regular expression into a comma separated string that was stored in a text field in the database as were counts of the number of GTPs and the number of unique Flickr photographer IDs. Next, the unique tags found within a given hexagon were sorted according to their frequency and the ten most frequently occurring tags were stored in another text field ("top10tags") as a comma separated string. Accompanying counts of occurrence of each tag ("freqtop10tags") were also stored for each hexagon in a similar manner. It should be noted that although the text tags that users record for GTPs are a potentially rich resource for analysis, their quality can vary substantially as many other forms of user-generated content does (Crandall et al., 2009; Graham & Zook, 2011, among others). For instance, users may record no tags on their photos, they may assign the same group of tags to many photos (i.e. bulk tagging such as "Summer 2012 vacation"), or may apply different terms or folksonomies in their classification of their photos. The average number of tags within our set of 54,522 GTPs was 9.8, however 4972 photos had no tags and one photo had 111 tags. The frequency distribution of tag counts for GTPs obtained for this study reveals a long-tailed distribution (Fig. 2).

Given that we were interested in comparing the occurrence of each of the top 10 tags in a hexagon relative to its frequency in its neighboring hexagons, the original PostGIS hexagon tables were restructured into new "count" tables. In the count tables, the compound "top10tags" and "freqtop10tags" fields were decomposed such that each row contained fields that listed a hexagon ID from the original hexagon tables, a single tag value and an associated "tagcount" value. These normalized table structures improved processing performance by avoiding the need to parse the comma separated text strings on-the-fly and permitted other fields pertaining to neighborhood tag occurrence counts to be processed easily. However, once the processing described in the next section was completed, the results were transferred to denormalized tables to ease cartographic symbolization and labeling.

### 2.3. Tag processing

Tag-similarity to its neighboring area was computed for each hexagon. Generally, we can anticipate that the way people characterize what they are photographing via their tag sets will be spatially autocorrelated, particularly for features or geographic "facts" for which there is little uncertainty or interpretation required (e.g. Niagara Falls, Eiffel Tower, etc.). Even in more nuanced contexts, we suggest that examining how the spatial pattern of tagging changes with location and hexagon size may provide insight into how people experience the urban environment. The difficulty of using traditional measures of association for categorical data is that in the map of tagsets, the number of possible tags changes across hexagons and across units of aggregation (hexagons sizes in this case). A simple measure of tag-set similarity based on the epidemiological concept of relative risk was initially used to compare the proportional abundances of tags in each hexagon relative to its neighbors (Lilienfeld, 1980), as:

$$TNrr_{ti} = \frac{\left[\frac{X_{ti}}{X_i}\right]}{\left[\frac{\sum_{j \neq i}^{k} X_{jt}}{\sum_{j \neq i}^{k} X_j}\right]} \tag{1}$$

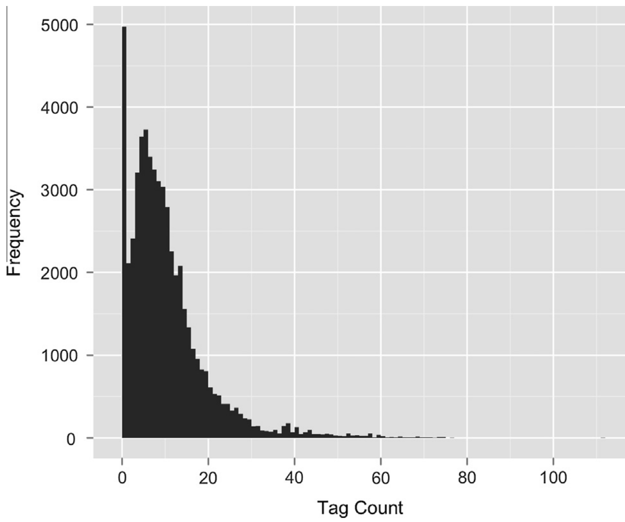


**Fig. 2.** Frequency distribution of tag counts from geotagged photos used in the study.

where $X$ is the frequency of a tag, indexed by $t$ which is an instance of one of the 10 most frequently occurring tags, and $i$ and $j$ are hexagon indices. As hexagons have symmetric neighborhood definitions, all non-edge hexagons had an equal number of neighbors (i.e. $k = 6$). In epidemiology, relative risk can be conceptualized as a basic $2 \times 2$ contingency table measure typically used to compare disease status in exposed and unexposed subpopulations. Within the context of this work, the measure above can be seen as analogous to the relative risk of each local hexagon's dominant tag being found in its neighborhood. It is essentially a measure of heterogeneity where values near 1 indicate similarity in proportional tagging, and extreme values indicate dissimilarity in tagging between hexagon $i$ and its neighboring hexagons. Since the TNrr as defined above is local to hexagon $i$ and tag $t$, aggregation of these measures are required across tags and hexagons. To aggregate across tags, we sum the TNrr for each hexagon across all of the top 10 tags. However because TNrr is asymmetric (i.e. bounded by zero and positive infinity with an expected value of one), we instead use the natural logarithm of the odds ratio of tag $t$ being in the neighborhood of hexagon $i$.

$$TNlor_{ti} = ln\left(\frac{\left(\left[\frac{X_{ti}}{X_i}\right]/1 - \left[\frac{X_{ti}}{X_i}\right]\right)}{\left(\left[\frac{\sum_{j\neq i}^{k} X_{jt}}{\sum_{j\neq i}^{k} X_j}\right]/1 - \left[\frac{\sum_{j\neq i}^{k} X_{jt}}{\sum_{j\neq i}^{k} X_j}\right]\right)}\right) \tag{2}$$

This alters the interpretability such that values around zero have similar neighborhood proportions for tag $t$. Values less than zero indicate that tag $t$ is more prevalent in their neighborhood, while those above zero have proportionally fewer instances of tag $t$ in their neighborhood. Ultimately, as we look at less frequent tags in this way, we can see how the pattern of tagging frequency varies spatially. We can then aggregate across tags, computing an average of non-zero *TNlor* values for all $t$.

$$TNlor_i = \frac{\sum_{t=1}^{10} X_t TNlor_{ti}}{X_t} \tag{3}$$

This has the impact of giving more weight to the odds ratios associated with the more frequently used tags (and thus more robust proportion estimates) hexagons (McNeil, 1996). We use this aggregate measure to examine patterns of tagging visually and across scales of analysis.

### 2.4. Multi-resolution analysis and visualization

Given that the underlying urban fabric is geographically variable and heterogeneous, an exploration of the GTP across varying units of aggregation was undertaken. Geographical analysis of landscapes requires specification of spatial sampling units and the extent of the study (i.e. grain and extent) – and the way in which GTPs are distributed across scales is an important marker of the underlying spatial processes. In the biogeography literature, geographical analysis has always emphasized explicit consideration of spatial scale, for example in the investigation of species-area relationships (Fortin & Dale, 2005; Fritz, Schuurman, Robertson, & Lear, 2013). We take a similar approach here to examine how the number of unique users taking photographs changes with the size of the analysis "neighborhood" or geographic unit. As individual users can dominate specific locations by uploading many photographs of the same area, many studies have used the number of unique users as a more robust measure of GTP frequency across space (Purves et al., 2011; Lee, Won, & McLeod, 2008). We examine the variability in the number of unique users per hexagon by computing the coefficient of variation (CV) of unique users for each hexagon size, which standardizes the sample standard deviation by the mean and allows comparison across sampling units of variable size. In this sense, maximum heterogeneity in unique users may suggest an appropriate scale for analysis and, potentially, more representative or reliable sampling by photos of a given place or location. This scale-centric use of the coefficient of variation can be seen to be complementary to tag-focused approaches Purves et al. (2011) and Hollenstein and Purves (2010) use to explore possible bias in the occurrence of specific tags owing to substantial differences in the number of photo postings users make. While it would be valuable to explore user variability both through tags and across scales, we focus on spatial representation only in this paper. In the following section, we contextualize our analysis by focusing on some key areas in Vancouver.

## 3. Results and discussion

As expected, the spatial distribution of GTPs is clustered in the main urban tourist and entertainment oriented core of the city (Fig. 3). As noted earlier, the total number of photos obtained for the study was 54,522. Within 1 km of a central downtown intersection (i.e. Burrard St & Robson St.) in downtown Vancouver, some 12,000 photos (22%) were found, while 41,054 photos (75%) were found within 5 km. The maps in Fig. 3 depict geometric five-class symbology, which is a classification scheme useful for visualizing heavily skewed data. The monocentricity of the GTP distribution becomes more varied at around the 16 ha hexagon scale, where more relatively high GTP hexagons are found in non-core locations.

The spatial distribution of the unique users can shed light onto how well different areas may be represented by GTPs based on the notion that locations with many different photographers are likely to be better sampled than locations where only a few people take many photographs (Lee et al., 2008). This follows research on other well-known forms of UGC such as OpenStreetMap (OSM) (Girres & Touya, 2010; Haklay, 2010), WikiMapia and Wikipedia (Kittur & Kraut, 2008; Niederer & van Dijck, 2010) where data quality has been found to improve non-linearly as contributors increase in number and their individual edits coalesce toward "true" values. However, this general linkage between user counts and data representativeness or accuracy may not always be as direct with GTPs. WikiMapia, OSM and similar projects that feature spatial UGC (e.g. Museum of Science FireFly Watch, geoBirds) are based on individuals' deliberate edits to shared "maps" or data products that

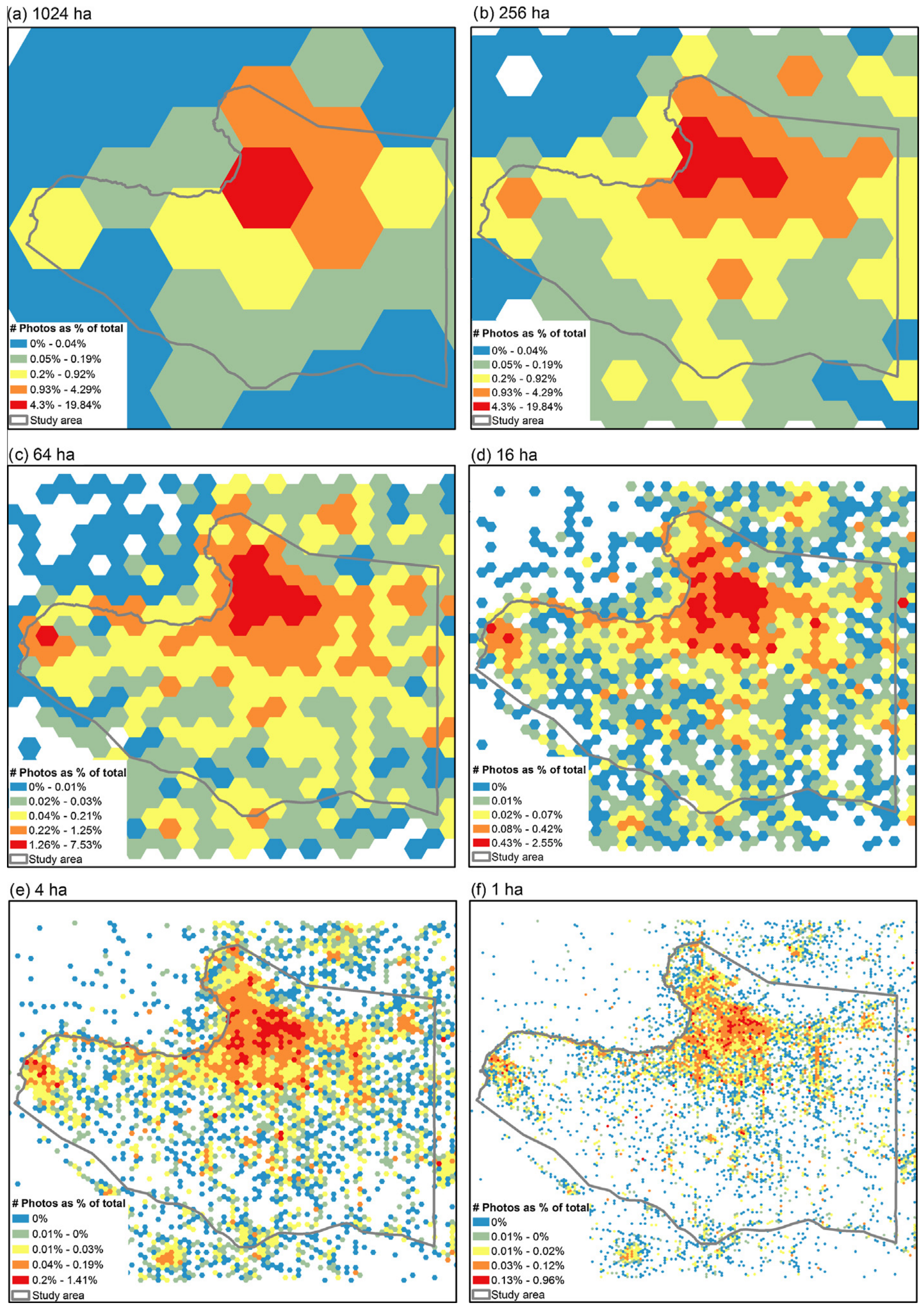


**Fig. 3.** Distribution of photos at 1024 ha (upper left) to 1 ha (lower right) resolutions.

focus on a limited set of tangible objects or "facts" (e.g. town boundaries, location of roads, names of building). In many cases, data collection is guided by pre-defined tag classes (e.g. OSM key-value pairs for road types, plant or animal specie lists). In contrast, Purves et al. (2011) and Stefanidis, Crooks, and Radzikowski (2013), among others, have noted that GTPs are a passive or even an ambient form of spatial UGC that people create and share for largely personal or social reasons, often without awareness of their potential as geographic data. Geographic and thematic coverage of GTPs, therefore, is less likely to be as consistent since individuals are free to choose what and where they photograph (e.g. friends, prominent landmarks, events, viewscapes), how their photos are composed (e.g. focal length, compass orientation) and how their photos are described through tags. Notwithstanding these caveats, it is reasonable to assume that areas with higher GTP user counts will be represented better, particularly if the area is characterized by a dominant landmark or shared socio-cultural interpretation (e.g. city hall, major sports arena, valued wilderness area).

In the Vancouver study area, the spatial pattern of unique users at larger hexagon sizes is fairly similar to that of the raw photo-count, however at smaller scales the patterns differed markedly. Fig. 4 presents unique user counts for the 1 ha hexagon scale in the downtown area. Areas of the city associated with tourism, recreation or entertainment (e.g. Gastown, Robson Square – near the "Downtown" label, Granville Island, BC Place, English Bay Beach) have high numbers of users, whereas most areas are captured by a relatively small subset of users. At first glance, it is somewhat curious that relatively few photographers were attracted to Stanley Park since the park is a well-known attraction for both residents and tourists. Two related factors can account for this. First, the park's size (400 hectares) coupled with the small hexagon size (1 ha or 107 m wide) contributes to lower user densities within individual hexagons. Second, photographers are much less confined in the park to defined corridors and nodes than elsewhere in the city and, with the exception of a few landmark attractions (e.g. Vancouver Aquarium, Brockton Lighthouse), photographers are more evenly dispersed. To illustrate the impact of urban morphology and especially transportation networks on GTP density, 27.5% of the Vancouver GTPs were found within 10 m of a road centerline, 50% were within 25 m and 66.4% were within 50 m.

These findings have important implications for small scale characterization of urban space using GTPs. If we examine the impact of scale explicitly on user counts, we see higher user counts associated with larger sampling units as may be expected. The southern and eastern portions of Stanley Park, for example, occupy mid-level classes when user counts are based on 16, 64 or 256 ha hexagons instead of the 1 ha units employed in Fig. 4. More interesting than user counts *per se*, is the dispersion in user counts. Dispersion (i.e. variance) in user counts indicates how evenly distributed users are across the study area. Typically dispersion is reduced as aggregation size increases, as local variability is smoothed over. This relationship is found in the Vancouver data when examining user count dispersion (measured by the coefficient of variation) versus hexagon size for all hexagons irrespective of whether any GTPs were recorded in them. However, if we focus only on areas that were sampled through GTPs and exclude hexagons with zero GTPs, Fig. 5 reveals increasing dispersion with hexagon size up to 16 ha, after which variation declines with increases in sampling unit size. This reflects the spatially "patchy" nature of GTPs which leaves a majority of small-size hexagons (i.e. 1 ha, .25 ha) with only one unique user, thereby reducing variation. It also suggests that spatial analysis of the Vancouver GTPs will not reveal generalizable patterns across the entire study area at scales smaller than 16 ha. However, as the following discussion reveals, smaller resolution hexagons can uncover interesting local features in areas with sufficient numbers of GTPs.

An example of the representativeness of GTPs at small scales is evident in the spatial discontinuity in unique user counts in Fig. 6. This figure centers on an area that transitions from tourist dominated urban landscapes in the west (Gastown, Chinatown) to an area of the city associated with poverty, drug use, and urban malaise in the center and east part of the map. The low GTP user counts at the .25 ha resolution in most of the "Downtown Eastside" are perhaps expected. However, it is interesting how well the .25 ha resolution reveals the sharp gradients in user counts between areas that are popular with residents and tourists and those that might be considered more typical aspects of the urban landscape and less likely to be photographed. For example, high unique user counts are found in the immediate vicinity of Simon Fraser University's (SFU) Vancouver campus (red hexagons in the upper left), along Water St. in the Gastown area, and the sports arenas to the south (Rogers Arena, BC Place). This demonstrates how spatial scale is conflated with urban use of space and some of the potential dangers in assuming that higher user numbers in applied GTP analysis necessarily translate to better data quality and/or more representative sampling of the urban environment.

Examining tag characteristics at multiple scales revealed significant dominance of a few select tags among the most frequently used tags (Table 2). The tags 'vancouver', 'Canada' and 'BC' were the only three tags among the top three tags across all scales. This finding corresponds to what others have found in similar analyses in other cities (e.g. Hollenstein & Purves, 2010). However, as hexagon sizes were reduced from 1024 ha to 1 ha, the proportional dominance of these top tags declined. For example, the top tag for 1024 ha represented 54% of all tags, while the top tag for 1 ha represented only 21% of all tags. This indicates that aggregation has a large impact on tag frequencies and thus the potential to extract useful information about locally discriminating place or event characteristics. Considered together with the changes in user frequency with scale and overall photocounts outlined earlier, it appears a balance must be sought in analysis of GTPs if both spatial and semantic content are to be analyzed jointly. We explored this

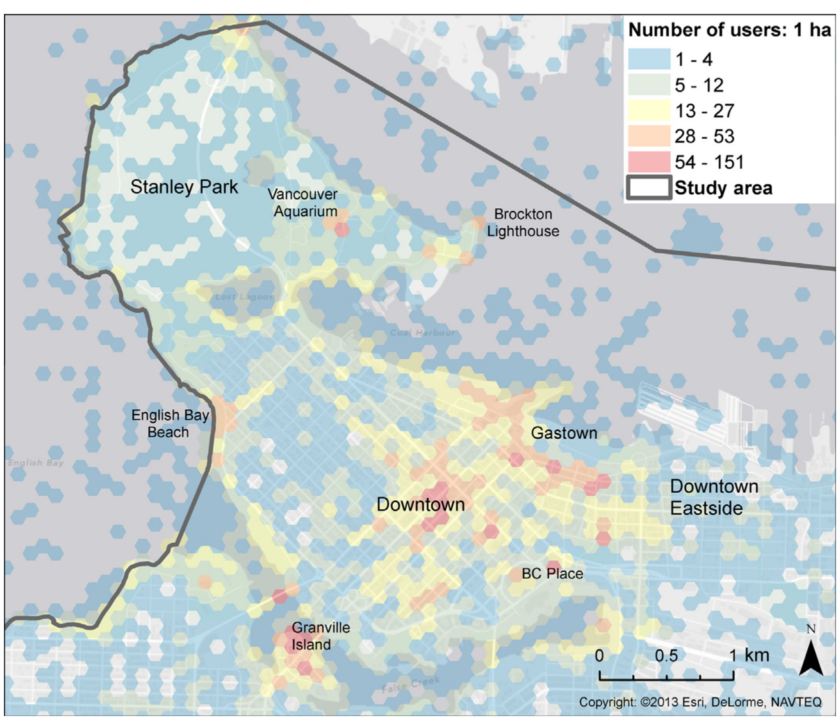


**Fig. 4.** Spatial distribution of unique users per hexagon at the 1 ha resolutions.

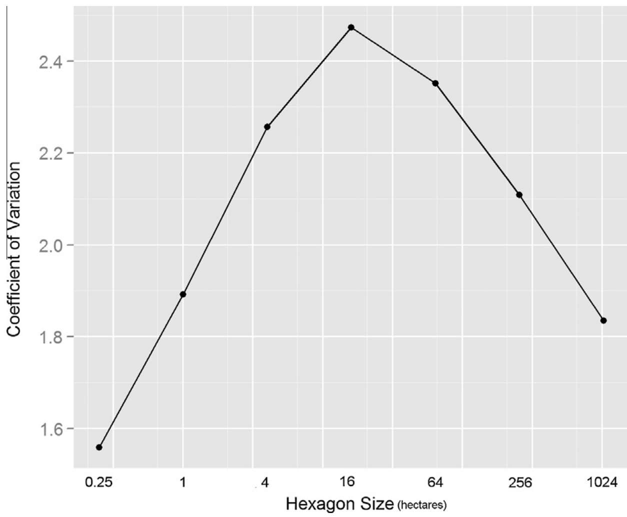


**Fig. 5.** Coefficient of variation of unique user count across multiple spatial scales.

idea further through the examination of neighborhood tagging characteristics.

Examination of tag similarity as measured by the neighborhood tagging log odds ratio (*TNlor*, see Eq. (3) in 3.2.2.) for multiple scales revealed that neighboring hexagons were more likely to have similar tags at smaller scales than for larger scales (Fig. 7). Given the two characteristics outlined earlier (i.e. tag dominance being greater at small scales, and user counts being very low at large scales), this is unsurprising. The increase in neighborhood tagging similarity is nonlinear, with a moderate increase between 1 ha and 16 ha, and larger increases thereafter. In this case, it appears that tag dominance overpowers any local (i.e. neighborhood) variance in proportional tagging at resolutions of 64 ha and larger.

To investigate the question of what may be the most appropriate scales to aggregate GTPs within urban contexts and to explore local spatial patterns in neighborhood tag similarity, the neighborhood odds ratios were simplified for map-based visualization. Recall that values of *TNlor* that are near zero indicate that a hexagon has similar tag characterization as its contiguous neighbors, while values above and below zero indicate that the phenomena described in the tags are, respectively, less and more likely to occur in the adjacent hexagons. The means (−.3856, −.3885, −.3447) and the standard deviations (.4109, .4376, .4388) of *TNlor* were quite consistent for 1 ha to 16 ha scales and reasonably close at the 64 ha scale (−.2689 and .4509) after which aggregation effects mute local tagging differences (−0.147466 and 0.506208 for 256 ha; −0.091899 and 0.557205 for 1024 ha).

To aid cross-scale visualization, we used a simple 3-category classification centerd on one-half of a standard deviation around the mean. In Fig. 8, this approach is used to explore tag-space at the 64 ha scale where tag-space similarity is highlighted in yellow, while areas that contributors have described using proportionally higher (red) or lower (blue) tag-combination frequencies relative to their neighbors. At this level of aggregation, there is a broad band of agreement in tag characterization throughout much of the central portion of the city (i.e. yellow hexagons in central business district and main tourist area). A considerable number of areas appear as relatively sparse in tag-space terms (blue), many of which can be traced to edge effects and infrequently photographed residential or industrial lands. The fewer red areas, which indicate localized uniqueness in tag-space through higher proportions of specific tag occurrences than their neighboring areas, appear to correspond to well-known and specific features or places (e.g. Lions Gate Bridge, Stanley Park, University of British Columbia – UBC).

More insight into the place-specific nature of users' tagging can be gained by examining the *TNlor* values calculated for smaller hexagons. Figs. 9–11 cover an area that includes the Canada Place cruise ship port on the north, the historic Gastown and Chinatown areas to the south and east and contrasts the 4, 1 and .25 ha resolutions. For reference, the locations of the Flickr photo data are shown as green dots. At the 4 ha resolution (Fig. 9), the red hexagons correspond to locally unique areas and features that are largely well-known and unambiguous to both residents and tourists. These areas have tags that occur with higher frequency relative to neighboring areas and include Canada Place and the nearby water taxi terminal in the upper left of the figure, Crab Park in the upper center and Rogers Arena, Chinatown and Dr. Sun Yat Sen Gardens in the south center. The blue hexagons throughout much of the Gastown tourism and entertainment district are also

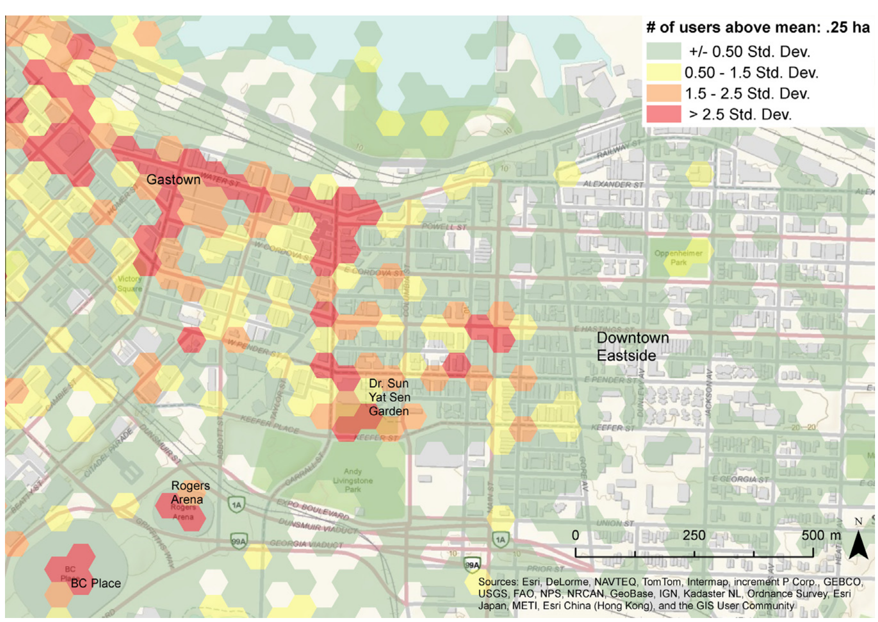


**Fig. 6.** Vancouver's Downtown Eastside – number of Flickr users with photographs in 1 ha sampling areas.

**Table 2**
Most frequently occurring tags in the dataset.

| Scale | Most frequent tag | 2nd most frequent tag | 3rd most frequent tag |
|---|---|---|---|
| 0.25 ha | Vancouver (1699, 16%) | BC (868, 8%) | Canada (844, 8%) |
| 1 ha | Vancouver (1296, 21%) | BC (556, 9%) | Canada (544, 9%) |
| 1024 ha | Vancouver (22, 54%) | Canada (11, 27%) | BC (12, 29%) |
| 16 ha | Vancouver (445, 33%) | BC (154, 11%) | Canada (155, 11%) |
| 256 ha | Vancouver (75, 54%) | Canada (28, 20%) | BC (31, 22%) |
| 4 ha | Vancouver (777, 24%) | BC (314, 10%) | Canada (300, 10%) |
| 64 ha | Vancouver (202, 45%) | Canada (70, 16%) | Canada (70, 16%) |

interesting since these low *TNlor* values are not due to low photo counts as described earlier for Fig. 8 at the 64 ha level. In this case, low *TNlor* values are indicative of a regional tag-space area that individuals are characterizing through more heterogeneous tagging relative to its neighbors.

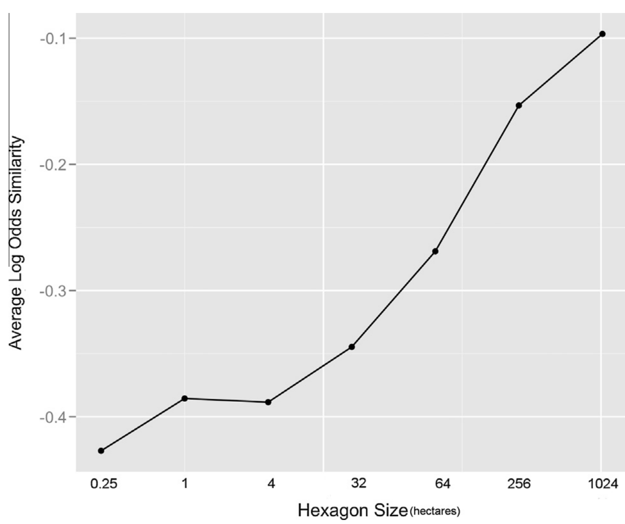


**Fig. 7.** Average log odds ratio of neighborhood similarity for 0.25 ha to 1024 resolutions.

As we move to smaller hexagon sizes, tag similarity becomes more varied. For example, when we compare the 4 ha (Fig. 9) and 1 ha (Fig. 10) resolutions, it is apparent that some areas with high *TNlor* values at the 4 ha scale (e.g. Canada Place, Crab Park, Rogers Place) continue to have high values at the 1 ha level, albeit with reduced areal extent. In contrast, several other areas (e.g. water taxi terminal south of Canada Place) change from high (red) to similar (yellow) tagging as their neighbors at the 1 ha resolution, while others are effectively split between two classes (e.g. red 4 ha hexagon to the west of the Downtown Eastside label). At the .25 ha resolution (Fig. 11), local variability of tagging is more apparent and fewer instances of high *TNlor* values are evident. While the dominance of .25 ha tag-space areas by neighborhood similarity (yellow) or proportionally low tagging frequencies (blue) is at least partially due to fewer photographs and photographers per hexagon, in some areas it may also be indicative of differential preferences, experiences and expressions of a common urban space through GTPs. This drilling down of tag-space through spatial scale offers an interesting lens through which to view the urban landscape. Further to the south at the center-bottom of the figure, Dr. Sun Yat Sen Chinese garden illustrates tag-space uniqueness and is shaded red which signals relative abundance. This specificity was not evident at the 64 ha scale of Fig. 8 above or at 64 ha (not shown).

At smaller spatial resolutions, the potential advantages of viewing urban GTPs through tag-space become more evident. As expected, the sequence of images in Figs. 9–11 shows greater heterogeneity with decreasing hexagon area. We suggest that the value of GTPs for urban analysis lies in this heterogeneity,

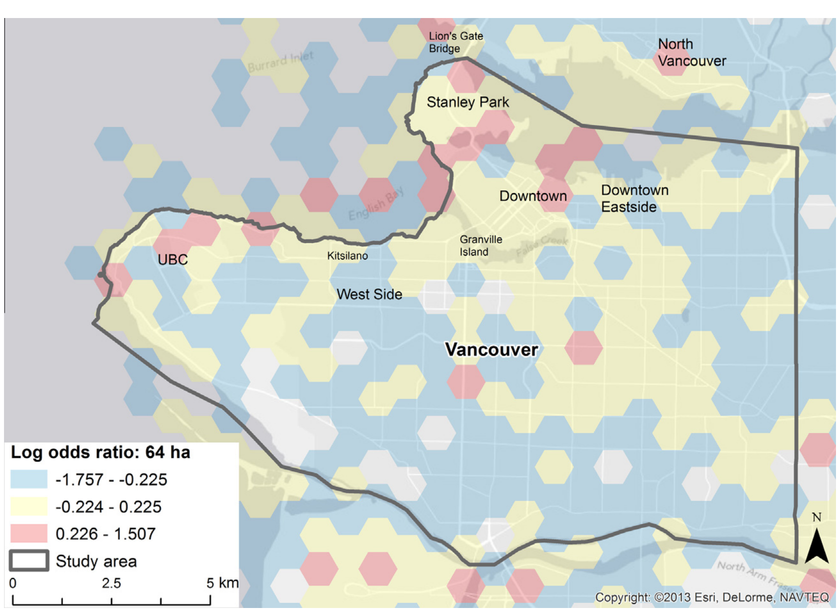


**Fig. 8.** Classified log odds ratio of neighborhood tag similarity at 64 ha.

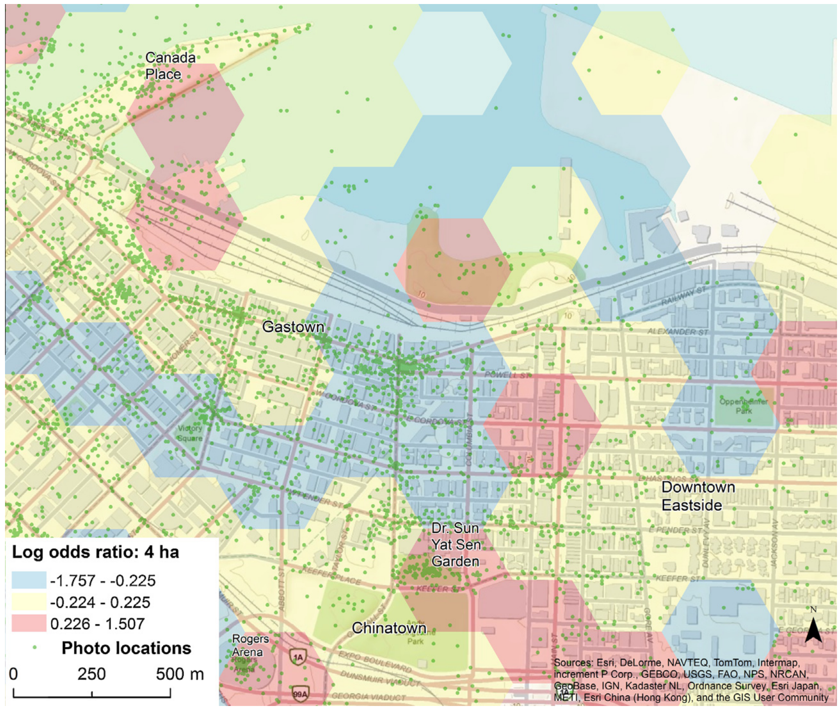


**Fig. 9.** Tagging similarity – 4 ha resolution.

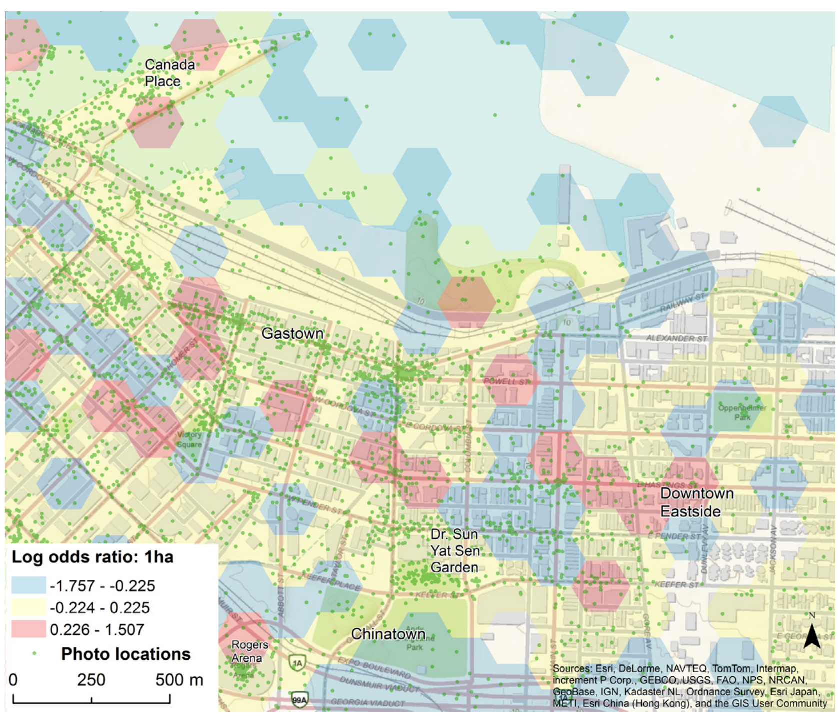


**Fig. 10.** Tagging similarity – 1 ha resolution.

especially in areas with higher photocounts and numbers of unique users where we may also be seeing, to some degree, the spatial definition of users' characterization of place. Comparing log odds ratios across different neighborhood sizes provides one avenue to begin unpacking the complexities of how places are defined in tag-space. Fig. 12 presents a multi-resolution map view of tag similarity across the 16 ha, 4 ha and 1 ha levels that aims to highlight the spatial variability in scale-dependent processes that are expressed in GTP tagging. In this figure, hexagon transparency is set proportional to sampling unit size such that larger hexagons are more transparent than smaller hexagons. Areas that have high tag similarity values at two or all three resolutions appear as shades of dark orange and red. This suggests local agreement across several partitions of tag-space. In contrast, the shades of blue and greens throughout the much of the figure characterize areas that have either sparse GTP counts and somewhat idiosyncratic tagging (e.g. blue-shaded residential neighborhoods in the south) or tagging that is largely dissimilar to neighboring areas

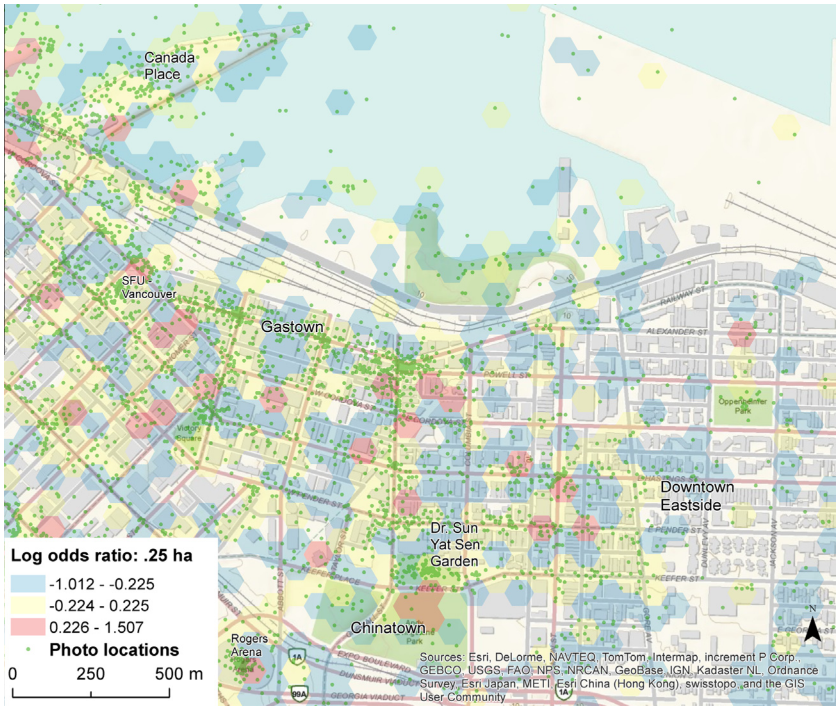


**Fig. 11.** Tagging similarity – 0.25 ha resolution.

(e.g. light and medium greens along many of the major roadways). Off-colors in this representation (i.e. shades of purple, light orange, etc.) indicate spatial disagreement across scales and offer perhaps a view into localized (i.e., geographically small) place-descriptors relative to their encompassing neighborhood. For example, Granville Island might be well-defined in tag-space at the 1 ha resolution, but at smaller scales more of the encompassing area that is not part of Granville Island leads to more varied representation in tag-space. Such scale-discontinuities may provide a means for automating localized place-delineations from tag-space representations.

However, as Fig. 13 below shows, the opposite occurs as well, where a well-defined place at a broader scale masks heterogeneity at a local scale. The main map is centerd on a region of agreement at the 4 ha resolution, located in the heart of Vancouver's most depressed neighborhood, known as the Downtown Eastside, an area that is characterized by lower incomes as well as drug use, prostitution, and homelessness (Smith, 2003; Quastel, 2009). In order to reveal greater granularity in tag-space and avoid the dominance of the most frequently recorded tags (see Table 2), the 5th to 8th most frequently occurring tags in each hexagon are stacked vertically as labels. At the 4 ha level, there are some tag references to local

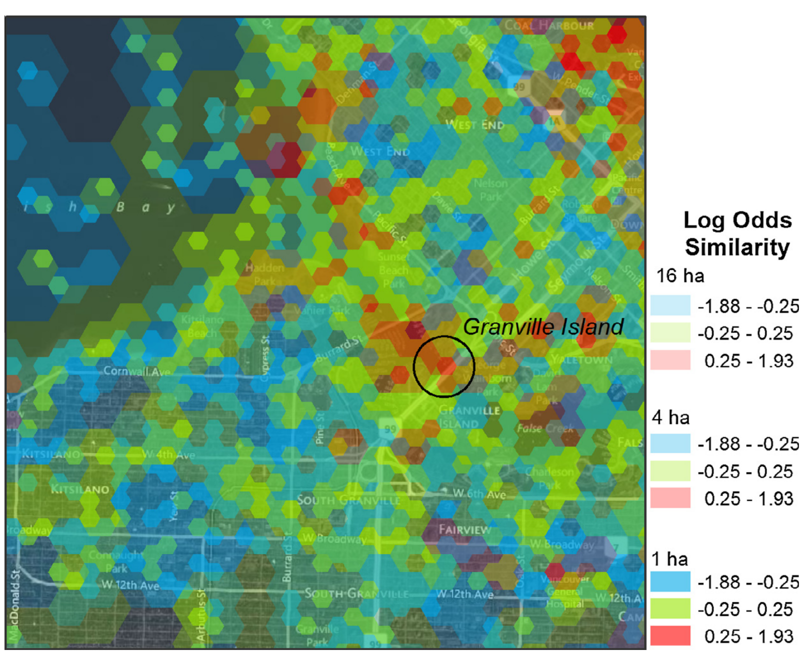


**Fig. 12.** Neighborhood tagging similarity across 16 ha, 4 ha and 1 ha hexagon sampling units – West End of downtown Vancouver.

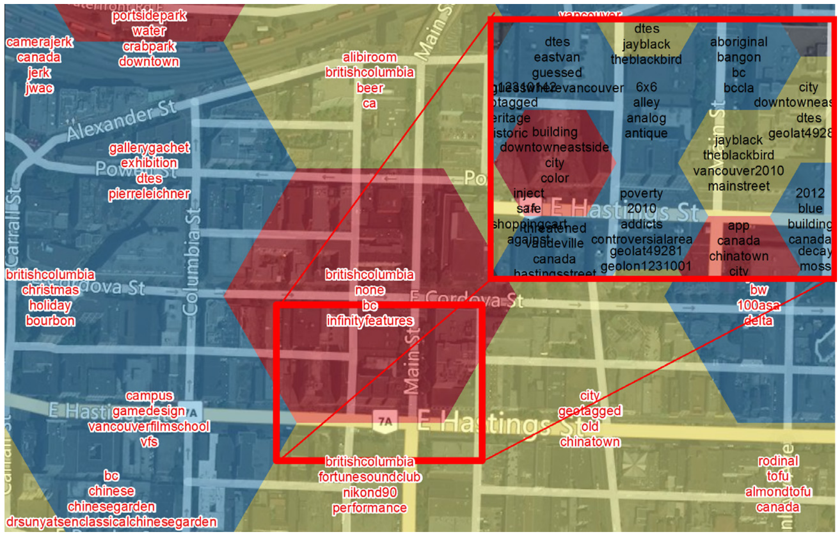


**Fig. 13.** False Creek area, Vancouver. The 5th through 8th most frequently used tags are shown as stacked labels for the 4 ha (main map) and 0.25 ha (inset map) resolutions. Neighborhood tagging similarity is indicated by yellow (similar) and blue or red (dissimilar). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

landmarks and art/entertainment businesses (e.g. "chinesegarden", "vancouverfilmschool", "fortunesoundclub", "alibiroom"). However, examining tags in the inset map reveals more disagreement and heterogeneity in tag-space. As context for viewing the tags at the .25 ha level, ongoing redevelopment and gentrification processes in the Downtown Eastside accelerated in advance of the 2010 Olympic Winter Games and led to increased concerns about the loss of affordable housing and the displacement of lower income populations (CBC News, 2012; Woo, 2013). Several of the tags in the inset .25 ha map highlight some aspects of this dynamic, in particular "controversialarea", "poverty", "addicts", "historic-building", and "jayblack" – an apparent reference to the Vancouver photographer who documented local residents' concerns and protests against the allocation of funds toward Olympic infrastructure through photo essays (Black, 2009). While this is a specific example, it provides some indications of how multi-resolution approaches can help unpack the often complex, overlapping and hierarchical nature of how people sense and represent urban environments through spatial UGC.

## 4. Conclusion

This paper demonstrates a method for analyzing tag frequency for a set of GTPs across multiple scales using Flickr data obtained for the city of Vancouver, Canada. In general terms, we demonstrated the dependence of tag characteristics on spatial scale of aggregation. At larger areas of aggregation, tag-space is dominated by a few frequent tags that describe large geographies, whereas more place-specific tags emerged at local scales. In the context of the Vancouver Flickr data, this effect was particularly evident for areas of 16 ha and lower. This was further highlighted by the equal dependence of neighborhood tagging with scales; as aggregation smooth out irregularities, neighboring sampling units are more similar in tag-space when examining frequently occurring tags.

Spatial patterns of GTPs revealed clustering in tourist and entertainment oriented areas of the city. A multi-scale perspective revealed significant spatial heterogeneity even when changing scales from 16 ha to 4 ha to 1 ha. The fractured spatialities evident in realized urban forms are highlighted by the degree to which tagging measures, in our case represented by the neighborhood log odds ratio, are altered when scales of aggregation change. This highlights a potential role for multi-scale approaches to help identify and delimit places that have complex, overlapping, and perhaps, as Graham, Zook, and Boulton (2012) suggest, contested meaning.

Further, as Dykes et al. (2008) and Purves et al. (2011) illustrate through their research using tree-maps to represent spatial and semantic similarity, our results suggest that scale effects should be considered explicitly as researchers explore the application of VGI for wider research and applied domains. As discussed above, Fig. 13 illustrates this potential where similarity at one scale is contrasted with heterogeneity at smaller scales – a clear example of the modifiable areal unit problem and the potential masking of scale-sensitive social processes when UGC/VGI are examined only at one level of aggregation. Note that the tags displayed in Fig. 13 are the 5th to 8th most frequently occurring at the 4 ha and .25 ha (inset map) levels. It is unclear how are measures of similarity demonstrated throughout this paper would change if we had focused only on these tag sets. As outlined in Table 2, the top three most frequently occurring tags are little value at the intra-city scales explored here although clearly they would be relevant if the study area was large enough to include neighboring municipalities in the broader metro-Vancouver region. More refined exploration of tag frequency thresholds should yield insight into the generality of these findings for GTPs and other forms of tagged UGC.

The analysis presented here is not without limitations or opportunities to further study the impact of scale, tag-space and meaning embedded in GTPs. First, there were significant temporal variations in GTP frequency and location during the study period that were beyond the scope of this article to explore. Extending our analysis to space–time may provide additional insights into how spatial expressions of urban experiences varies in response to changes in the underlying physical "placescape" and more ephemeral personal or community events. The Winter Olympics in February of 2010 and the riots that followed the Vancouver Canucks 2011 Stanley Cup final loss are two events that appear promising in preliminary examination of the Vancouver data. In both cases, these events were associated with large influxes of visitors and residents to specific locales in the city which may provide interesting opportunities to explore any deviations from what might be considered the 'normal' pattern of GTP activity patterns.

Second, as our research sought to understand spatial forms and processes better at several urban scales as depicted through Flickr

UGC, establishing what can be considered 'local' versus 'non-local' required some spatial definition. Virtually all spatial analysis methods require a neighborhood definitions of some form (e.g. distance band, contiguity, etc.) that are either pre-defined spatial units or are calculated based on the spatial distribution of the data. Our approach of aggregating the Flickr data at several fixed resolutions draws upon the rich tradition of multi-scale analyses in the geography, ecology and epidemiology literature (see, for example, Fortin & Dale, 2005) and enables us to explore local and non-local spatial relationships explicitly through the lens of space and tag-space interactions. Fixed tessellations provide a stable framework to examine changes over time to GTP density and the ways that people characterize places. However, alternative approaches such as kernel density estimation and spatial scan methods (see, for example, Crandall et al., 2009; Fritz et al., 2013; Hollenstein & Purves, 2010; Rattenbury & Naaman, 2009) allow GTP clusters or regions of interest to be defined in a data sensitive and anisotropic manner. Since urban GTP locations are influenced strongly by road network and open space configurations (see Fig. 10), these methods appear particularly well-suited for detecting and delimiting GTP-sourced places that may be represented by several adjacent hexagons in the method illustrated in this paper.

Finally, the results reported in this research are confined to a single city and should be viewed as a first step towards development of scale-sensitive metrics of urban place in GTPs. In the current study, the authors' knowledge of the study site greatly aided interpretation of the patterns of GTPs across space- and tag-space. We intend to apply these methods to a wider variety of urban settings through an experimental control study design. Future research should aim to further generalize these space-scale-place relationships through more objective, automated implementations across more varied and larger study areas.

## Acknowledgements

The authors would like to thank the Social Sciences and Humanities Research Council (SSHRC) of Canada for supporting this research. In addition, the authors are grateful to the three anonymous reviewers for insightful suggestions that have substantially improved this article.

## References


Andrienko, N., & Andrienko, G. (2011). Spatial generalization and aggregation of massive movement data. *IEEE Transactions on Visualization and Computer Graphics, 17*(2), 205–219.

Antoniou, V., Morley, J., & Haklay, M. (2010). Web 2.0 geotagged photos: Assessing the spatial dimension of the phenomenon. *Geomatica, 64*(1), 99–110.

Beasley, L. (2000). Living first in downtown Vancouver. *Zoning News*, 1–4.

Black, J. (2009, October 16). A two-year march against homelessness. *The Tyee*. Retrieved from <http://thetyee.ca/News/2009/10/16/RenovictionCity2/>. Accessed 26.09.13.

Boots, B., & Tiefelsdorf, M. (2000). Global and local spatial autocorrelation in bounded regular tessellations. *Journal of Geographical Systems, 2*, 319–348.

CBC News (2012, December 26). "*Vancouver's Downtown Eastside feeling gentrification squeeze" - British Columbia - CBC News*. Retrieved from <http://www.cbc.ca/news/canada/british-columbia/vancouver-s-downtown-eastside-feeling-gentrification-squeeze-1.1190127>. Accessed 26.09.13.

Chen, R., & Xie, J. (2008). Open source databases and their spatial extensions. In G. B. Hall & M. Leahy (Eds.), *Open source approaches in spatial data handling* (pp. 105–129). Springer-Verlag.

Coleman, D. J., Georgiadou, Y., & Labonté, J. (2009). Volunteered geographic information: The nature and motivation of produsers. *International Journal of Spatial Data Infrastructure Research, 4*, 332–358.

Connors, J. P., Lei, S., & Kelly, M. (2011). Citizen science in the age of neogeography: Utilizing volunteered geographic information for environmental monitoring. *Annals of the Association of American Geographers, 102*(6), 1267–1289.

Crandall, D. J., Backstrom, L., Huttenlocher, D. & Kleinberg, J. (2009). Mapping the world's photos. In *Proceedings of the 18th international conference on world wide web* (pp. 761–770). ACM.

Crandall, D., & Snavely, N. (2012). Modeling people and places with internet photo collections. *Communications of the ACM, 55*(6), 52–60.

De Choudhury, M., Feldman, M., Amer-Yahia, S., Golbandi, N., Lempel, R. & Yu, C. (2010). Automatic construction of travel itineraries using social breadcrumbs. In *Proceedings of the 21st ACM conference on hypertext and hypermedia* (pp. 35–44). ACM.

Dykes, J., Purves, R., Edwardes, A., & Wood, J. (2008). *Exploring volunteered geographic information to describe place: Visualization of the 'Geograph British Isles', GISRUK 2008*. England: Manchester (pp. 256–267). England: Manchester.

Economist Intelligence Unit (2012). A summary of the liveability ranking and overview. <http://www.eiu.com/site_info.asp?info_name=The_Global_Liveability_Report>. Retrieved 01.02.13.

Elwood, S., Goodchild, M. F., & Sui, D. (2012). Researching volunteered geographic information: Spatial data, geographic research and new social practice. *Annals of the Association of American Geographers, 102*(3), 571–590.

Feick, R., & Roche, S. (2012). Understanding the value of VGI. In D. Sui, S. Elwood, & M. Goodchild (Eds.), *Crowdsourcing geographic knowledge: Volunteered Geographic Information (VGI) in theory and practice* (pp. 15–29). Springer.

Ferarri, L., Rosi, A., Mamei, M. & Zambonelli, F. (2011). Extracting urban patterns from location-based social networks. In *Proceedings of the 3rd ACM SIGSPATIAL international workshop on location-based social networks LBSN '11* (pp. 9–16). New York, NY, USA: ACM. doi:10.1145/2063212.2063226.

Fortin, M. J., & Dale, M. R. T. (2005). *Spatial analysis: A guide for ecologists*. Cambridge University Press.

Fritz, C. E., Schuurman, N., Robertson, C., & Lear, S. (2013). A scoping review of spatial cluster analysis techniques for point-event data. *Geospatial Health, 7*, 183–198.

Girres, J.-F., & Touya, G. (2010). Quality assessment of the French OpenStreetMap dataset. *Transactions in GIS, 14*(4), 435–459.

Goodchild, M. F. (2007). Citizens as sensors: The world of volunteered geography. *GeoJournal, 69*(4), 211–221.

Graham, M., Zook, M. & Boulton, A. (2012). Augmented reality in urban places: Contested content and the duplicity of code. In *Transactions of the institute of British geographers*. Article first published online 10.08.12, DOI: 10.1111/j.1475-5661.2012.00539.x.

Graham, M., & Zook, M. (2011). Visualizing global cyberscapes: Mapping user-generated placemarks. *Journal of Urban Technology, 18*(1), 115–132.

Haklay, M. (2010). How good is volunteered geographic information? A comparative study of OpenStreetMap and Ordnance Survey datasets. *Environment and Planning B, 37*, 682–703.

Hermida, A. (2010). Twittering the news. *Journalism Practice, 4*(3), 297–308.

Hollenstein, L., & Purves, R. (2010). Exploring place through user-generated content: Using Flickr tags to describe city cores. *Journal of Spatial Information Science, 1*, 21–48.

Jankowski, P., Andrienko, N., Andrienko, G., & Kisilevich, S. (2010). Discovering landmark preferences and movement patterns from photo postings. *Transactions in GIS, 14*(6), 833–852.

Jin, X., Gallagher, A., Cao, L., Luo, J. & Han, J. (2010). The wisdom of social multimedia: Using Flickr for prediction and forecast. In *Proceedings of the international conference on multimedia (MM '10)* (pp. 1235–1244). ACM.

Kear, M. (2007). Spaces of transition spaces of tomorrow: Making a sustainable future in Southeast False Creek, Vancouver. *Cities, 24*(4), 324–334.

Kennedy, L., Naaman, M., Ahern, S., Nair, R. & Rattenbury, T. (2007). How Flickr helps us make sense of the world: Context and content in community-contributed media collections. In *Proceedings of the 15th international conference on multimedia MULTIMEDIA '07* (pp. 631–640). New York, NY, USA: ACM.

Kisilevich, S., Keim, D., & Rokach, L. (2010). A novel approach to mining travel sequences using collections of geotagged photos. In M. Painho, M. Santos, & H. Pundt (Eds.), *Geospatial, thinking* (pp. 163–182). Springer.

Kittur, A. & Kraut, R.E. (2008). Harnessing the wisdom of crowds in wikipedia: Quality through coordination. In *Proceedings of the 2008 ACM conference on computer supported cooperative work (CSCW '08)* (pp. 37–46). New York, NY, USA: ACM. DOI: 10.1145/1460563.1460572.

Lee, S., Won D., & McLeod, D. (2008). Tag-geotag correlation in social networks. In *Proceedings of the 2008 ACM workshop on search in social media (SSM '08)* (pp. 59–66). New York, NY, USA: ACM. http://dx.doi.org/10.1145/1458583.1458595.

Lilienfeld, A. M. (1980). *Foundations of epidemiology* (2nd ed). New York: Oxford University Press.

McKenzie, P., Burkell, J., Wong, L., Whippey, C., Trosow, S., & McNally, M. (2012). User-generated online content 1: Overview, current state and context. *First Monday, 17*(6). http://dx.doi.org/10.5210/fm.v17i6.3912.

McNeil, D. (1996). *Epidemiological research methods*. New York: J. Wiley.

Moxley, E., Kleban, J. & Manjunath, B.S. (2008). Spirittagger: A geo-aware tag suggestion tool mined from Flickr. In *Proceeding of the 1st ACM international conference on multimedia, information retrieval* (pp. 24–30). ACM.

Niederer, S., & van Dijck, J. (2010). Wisdom of the crowd or technicity of content? Wikipedia as a sociotechnical system. *New Media & Society, 12*(8), 1368–1387.

Patil, G. P., Myers, W. L., Luo, Zhen, Johnson, G. D., & Taillie, C. (2000). Multiscale assessment of landscapes and watersheds with synoptic multivariate spatial data in environmental and ecological statistics. *Mathematical and Computer Modelling, 32*(1–2), 257–272. http://dx.doi.org/10.1016/S0895-7177(00)00133-3.

Purves, R., Edwardes, A. & Wood, J. (2011). Describing place through user generated content. *First Monday 16*(9), (September 5). <http://openaccess.city.ac.uk/535/> Retrieved 01.02.13, http://dx.doi.org/10.5210/fm.v16i9.3710.

Quastel, N. (2009). Political ecologies of gentrification. *Urban Geography*. http://dx.doi.org/10.2747/0272-3638.30.7.694.

Rattenbury, T., & Naaman, M. (2009). Methods for extracting place semantics from Flickr tags. *ACM Transactions on the Web, 3*(1), 1–30.

Rempel, R.S., Kaukinen, D., & Carr, A.P. (2012). *Patch Analyst and Patch Grid. Ontario Ministry of Natural Resources*. Centre for Northern Forest Ecosystem Research, Thunder Bay, Ontario.

Rorissa, A. (2010). A comparative study of Flickr tags and index terms in a general image collection. *Journal of the American Society for Information Science and Technology, 61*(11), 2230–2242.

Sheppard, E., & McMaster, R. (2004). *Scale and geographic inquiry: Nature, society, and method*. Oxford: Blackwell.

Shirkey, C. (2008). *Here comes everybody: The power of organizing without organizations*. New York: Penguin.

Smith, H. (2003). Planning, policy and polarisation in Vancouver's downtown eastside. *Tijdschrift voor economische en sociale geografie, 94*(4), 1467–9663.

Stefanidis, A., Crooks, A., & Radzikowski, J. (2013). Harvesting ambient geospatial information from social media feeds. *GeoJournal, 78*, 319–338.

Sui, D., & Goodchild, M. F. (2011). The convergence of GIS and social media: Challenges for GIScience. *International Journal of Geographical Information Science, 25*(11), 1737–1748.

Tapscott, D., & Williams, A. (2006). *Wikinomics: How mass collaboration changes everything*. New York: Portfolio.

Turner, M. G., Gardner, R. H., & O'Neill, R. V. (2001). *Landscape ecology in theory and practice: Pattern and process*. Springer.

White, D., Kimerling, J. A., & Overton, S. W. (1992). Cartographic and geometric components of a global sampling design for environmental monitoring. *Cartography and Geographic Information Science, 19*(1), 5–22. http://dx.doi.org/10.1559/152304092783786636.

Wiersma, Y. F. (2010). Birding 2.0: Citizen science and effective monitoring in the Web 2.0 world. *Avian Conservation and Ecology 5*(2). <http://dx.doi.org/10.5751/ACE-00427-050213> Retrieved 04.05.13.

Wilson, R. E., Gosling, S. D., & Graham, L. T. (2012). A review of Facebook research in the social sciences. *Perspectives on Psychological Science, 7*(3), 203–220.

Woo, A. (2013, August 22). Activists in Vancouver's Downtown Eastside challenge latest attempt to gentrify neighbourhood. *The Globe and Mail*. <http://www.theglobeandmail.com/news/british-columbia/residents-of-vancouvers-downtown-eastside/article13930493/> Accessed 26.09.13.

Wu, L., Hua, X., Yu, N., Ma, W. & Li, S. (2008). Flickr distance. In *Proceedings of the 16th ACM international conference on multimedia MM '08* (pp. 31–40). New York, NY, USA: ACM.

Zhang, H., Korayem, M., You, E. & Crandall, D. J. (2012). Beyond co-occurrence: Discovering and visualizing tag relationships from geo-spatial and temporal similarities. In *Proceedings of the fifth ACM international conference on web search and data mining (WSDM'12)* (pp. 33–42). Seattle, Washington, USA: ACM.