The struggle to belong

Dealing with diversity in 21st century urban settings.

Amsterdam, 7-9 July 2011

Exploring urban place: the potential, challenges and limitations of Volunteered Geographic Information

Ate Poorthuis and Karin Pfeffer (*)

Paper presented at the International RC21 conference 2011
Session: **RT1.2: Mixed methods.**

(*) University of Amsterdam
Nieuwe Prinsengracht 130
1018VZ Amsterdam, The Netherlands
a.poorthuis@uva.nl // k.pfeffer@uva.nl

Abstract

The data sources that researchers, policy makers and practitioners rely upon are becoming increasingly diversified and dynamic. Such data is no longer necessarily produced by traditional 'experts' (e.g. government), because today, laymen are also able to participate in data creation, consumption and exchange (Goodchild, 2007). This trend is of particular interest with regards to geographic data, as the combination of user-friendly software for geographical information analysis and location-based technology (often integrated in mobile phones etc.) has created an extensive infrastructure for user-created and place-specific data. This type of data is often labeled Volunteered Geographic Information (VGI). OpenStreetMap (www.openstreetmap.org) is one example of a site where a wide range of spatial data can be created, edited and viewed by anyone.

In this this paper we examined critically the potential, challenges and limitations of VGI for social science research. We used approximately 200,000 Flickr.com photographs and their descriptions to analyze the dynamics of place in Amsterdam. The study looked at the location, meaning and heterogeneity of *place* – showing how tourists and local inhabitants perceive the city in different ways. Apart from the potential, we looked at the challenges and limitations of using VGI data, mainly around representation, distribution and aggregation. The results show that specific groups of people are under- or overrepresented. In addition, VGI is not evenly distributed across geographic space. Both popular and affluent areas receive a disproportionate amount of attention. Finally, since place – fuzzy and fluid – does not exist only on a specific scale, issues of aggregation arise when interpreting the rigid spatial data (points, lines, polygons) derived from VGI. Our analysis shows that, on one hand, VGI can be a rich data source for social science research. On the other hand, data quality and control are pressing issues that need further research.

Introduction

The commodification of new information technologies (IT) in the last 10 to 15 years has led to a stage in which social scientists (amongst others) cannot only *use* information technology as a tool to further their research but are also being challenged to *(re-)visit* classical research strands using the new datasets that have been produced with these new technologies[1].

In recent years we have seen a whole range of new technologies and online services that have been rapidly changing geographic information as such. Individuals – lay people - are now using the Web to create, use, obtain, and disseminate geographic information thanks to sites such as Wikimapia and OpenStreetMap, while Google Maps, Microsoft's Visual Earth and Google Earth enable users to develop their own applications, analyze spatial data and disseminate the results in a relatively easy manner. Not surprisingly, the developments in this field have become a subject in a wide range of academic research. From the geographical perspective, especially interesting in this light

[1] In addition, information technology is increasingly the subject of research itself (cf. Kitchin and

is the emergence of neogeography or volunteered geographic information (VGI), a rather new research field concerning user-generated geographic information.

In his influential article 'Citizens as sensors: the world of volunteered geography,' Goodchild (2007) relates this new phenomenon to the traditional creation of geographic knowledge. Traditionally this production is a top-down governed process, often state-led, in which only 'experts' and 'professionals' participate: *"(…) large numbers of private citizens, often with little in the way of formal qualifications, [engage] in the creation of geographic information, a function that for centuries has been reserved to official agencies. They are largely untrained and their actions are almost always voluntary, and the results may or may not be accurate. But collectively, they represent a dramatic innovation that will certainly have profound impacts on geographic information systems (GIS) and more generally on the discipline of geography and its relationship to the general public. I term this volunteered geographic information (VGI).*" (Goodchild, 2007). No longer is the production of (geographic) knowledge the exclusive domain of the state (Scott, 1998).

While the basis was laid down by designers (cf. Stamen Design[2]) and computer scientists (cf. Yahoo Research Lab[3]), social scientists are recognizing the potential these tools and data have for developing new research questions and directions. While it seems to be easier to publish research on subjects when they just emerge in the academic arena (c.f. the number of scientific publications concerning new IT and social media in the last three years), it is much more difficult to substantiate the scientific relevance these subjects have. Since we seem to have moved past the height of the hype around new IT and social media, it is time to take a step back and re-evaluate the use and value of VGI. Is it just 'hot air' or do these new tools and data provide sufficient substance? Underlying the enthusiasm and excitement about the possibilities and opportunities of IT and VGI, there is a much more fundamental debate which is loosely centered around the potential benefits (and limitations) of globalization and its motor: the communication (r)evolution. On the one hand, some believe that this evolution is flattening the world: geographical differences become smaller and might even disappear over time (cf. Fukuyama, 1992). This is a popular strand of thought that has high-profile advocates, for example the oft-cited Friedman (2006). Already a decade before Friedman, Cairncross (1997) made a similar claim: "*In all these ways, the communication revolution is profoundly democratic and liberating, leveling the imbalance between large and small, rich and poor. The death of distance, overall, should be welcomed and enjoyed. "*

On the other hand, however, geographers have shown otherwise for those that are affected by these differences and uneven geographies. Most critical geographers seem to agree that -- on the contrary -- globalization has intensified differences among space, place and people (Brenner, 2008; Dodge and Kitchin, 2001; Graham, 2002; Warf, 2001). Parallel to Cairncross, Castells (1997) is pinpointing the problematic issues of the information and communication technologies.

---

[2] http://stamen.com/

[3] http://research.yahoo.com/

"While a relatively small, educated, and affluent elite in a few countries and cities would have access to an extraordinary tool of information and political participation, actually enhancing citizenship, the uneducated, switched off masses of the world, and of the country, would remain excluded from the new democratic core, as were slaves and barbarians at the onset of democracy in classical Greece. "(Castells, 1997)

There is a (growing) gap between those that have the financial and physical resources and necessary skills to use new communication technology and those who do not have that access, commonly known as the 'digital divide'. So new (communication) technologies open up a wide range of opportunities for a small number of privileged, while neglecting the disadvantaged.

The purpose of this article is to contribute to the critical view on VGI that has been initiated in the beginning of the 21st century by a number of geographers (cf. Sui, 2008; Zook, 2004; Purves and Edwardes, 2008; Elwood, 2010). The particular aim is to investigate the potential, challenge and limitations of VGI for studying a typical human geography concept, namely the perception of place. We will do that by analyzing how urban place, and in particular the city of Amsterdam, is consumed and produced through the photo-sharing website Flickr.com. The structure of the article is the following. First we will introduce our conceptualization of place and spend a few words on cyberspace and the digital divide. Then we will explain the research methodology and methods, including a description of Flickr.com and the dataset derived from that platform. This is followed by the analytical part of this paper, in which we look at several dimensions of place as well as the uneven geography ('divide') within our dataset. Finally, we will discuss the potential, challenges and limitations of VGI within the context of social sciences, and for studying place in particular.

Place

Place has to be one of the most multi-layered and multipurpose keywords in our language---Harvey (1996)

In its core, human geography is about the study of places (Cresswell, 2004). However, despite the central role that place has within the discipline, its meaning remains deeply contested, vague and abstract[4]. Academically, the challenge lies in trying to define something that is not tangible, a word so physical while without physical substance and perhaps most significantly, defined boundaries. The difficulty in teasing out a common meaning is exacerbated by the fact that the notion of 'place' is on the one hand highly academic, while at the same time commonly and often informally used in everyday language. I can invite you to come to *my place*. You can think of your home city as a *nice place*. A person can move to a new location, then settle and find his or her *place*. That sounds deceivingly basic, but it is obvious that place is

[4] See for example Sack's work on place (1997) and the amusing critique written by Symanski (2002)

a multi-faceted concept also in everyday language. It includes ownership, values, position and a feeling of connection. Unfortunately, the debate often remains rather abstract and philosophical. Seldom does it result in insights that can be readily used in applied research. For our conceptualization of place, we draw heavily on Tuan (1977), Taylor (1999), Harvey (1996) and Massey (1994). Combining the insights of these thinkers, we distinguish six dimensions of place:

1. Place has a location. In other words, place is somewhere on the earth's surface. This can be exact, like the geographic coordinates of my university library. It can also be a geographic reference only, like New York City. We neither know exactly where the center of New York City is nor what its extent is. But that is not so important. All that matters is that it exists somewhere.
2. Place has meaning. This is the 'sense of place' that so many authors are referring to. Since place is something personal that humans are attached to, it becomes meaningful. It is a subjective, emotional meaning. *My* university library has a specific meaning to me -- it can be a safe, quiet, inspiring place but it might also be a place of horror and anxiety for *others*.
3. Place is heterogeneous. Considering the subjective, individual meaning of place, we derive that place must be heterogeneous. Places do not have singular identities. These identities might be conflicting and contested or simply co-existing.
4. Place is dynamic (through time). Place is not a pause in an otherwise mobile space. Place itself is fluid. It is continuously changing. The university library can have a different meaning to me now than it will have after graduation. A city park in the summer has a different meaning than the very same park in winter.
5. Place has fluid borders. This relates very much to the previous points. It is often impossible to pinpoint an exact location of place. While sometimes the centre of that place can be gauged, determining the geographic boundaries of a place is difficult. These differ from person to person, from meaning to meaning and from time to time. For some, Staten Island might not be part of New York City while for others Staten Island may constitute the very heart of New York City.
6. Place is socially constructed. It does not exist 'naturally' but is actively constructed by society and its individual members. That means individual people also have agency when it comes to how they construct place.

In this paper we will limit our analysis of place mainly to the first three dimensions.

Cyberspace and the Digital Divide

The important role information technologies play in globalization also leads to the discursive creation of 'cyberspace': 'a multi-media skein of digital networks which is infusing rapidly into social, cultural and economic life' (Graham, 1998, p. 165). The exact meaning of this new concept is widely discussed and disagreed on. In popular culture, it is often seen as something that is truly ‘virtual’ and might even replace territoriality and ‘real’ place. However, in practice -- as is

very clear in this paper -- cyberspace and 'real' place are interlinked in a myriad of ways.

It is important here to address a misconception about cyberspace that is drawn upon the technological determinism that is prevalent in popular media: the internet, or cyberspace, is the great leveler. It is supposed to be inherently democratic (cf. Cairncross, 1997). Following Warf (2001), we do not support these assumptions and consider them as unrealistic hopes and fantasies. If anything, we suggest to focus research and policy especially on the digital divide and its implications. In this paper, this notion of the uneven geography of cyberspace is a key issue. In fact, just as space itself, cyberspace is not free of power relations and hierarchies (Sheppard, 2002). It does not provide an opportunity to level the playing field, as is often thought, since access to cyberspace is not uniformly distributed. It is unable to free itself of social, geographical and historical constraints because new technologies are heavily embedded exactly within society, geography and history. This is nothing new and has been similar with the advent of other new communication technologies like telegraphy, radio and telephony (Hugill, 1999; Perkins and Neumayer, 2009).

Like other telecommunications systems, the Internet is a social product, interwoven with relations of class, race, and gender, and inescapably subject to the uses and misuses of power. Telecommunications are not inherently emancipatory, freeing people from `the tyranny of distance', as they can be used to monitor everyday life, including credit cards, visas and passports, tax records, medical data, police reports, telephone calls, utility records, automobile registration, crime statistics, and sales receipts (Warf, 2001)

The issue of access to ICTs the associated digital divide is important because it has far-fletching consequences. Those without access or with restricted access might very well be (or become) the have-nots of today's society (Mitchell, 1995).

Case of Amsterdam

This paper is based on comparative research comparing Amsterdam and New York City. However, in this paper we will focus solely on the case of Amsterdam. A city with nearly 800,000 inhabitants[5], Amsterdam is a diverse city. It is home to the rich and famous of the Netherlands but a sizable part of its residents are deprived. Similarly, the city is ethnically diverse, housing large groups of Turks, Moroccans, Antilleans and Surinamese. Only about half of the population is 'autochthon' (both parents born in the Netherlands). Not only is Amsterdam a diverse city, it is also a segregated city: the unemployment rate in 2010 was 5.2% in the Centre borough, while standing at 9.4% in the Zuidoost borough[6]. Of interest here is that the broadband penetration rate is much higher than in other major urban areas – about 83% in the Netherlands and reportedly even higher in

[5] http://www.os.amsterdam.nl/pdf/2010_kerncijfers_amsterdam

[6] ibid.

Amsterdam[7]. In the light of the prerequisite internet access for most VGI participation, that makes Amsterdam quite an interesting case.

Data, Methods and Approach

To analyze the perception of place, we use metadata of photographs that have been uploaded to the photo-sharing platform Flickr.com between 2004 and December 2007 and are located somewhere in the city of Amsterdam. The metadata of the photographs is processed by means of a geographical information system (GIS) to analyze the various dimensions of place as listed above. In the following section we will first turn to Flickr before we will go into the details of data mining and processing.

What is Flickr?

In the early 2000s the Flickr.com web application was developed by a small Canadian company that was founded by Stewart Butterfield, Caterina Fake and Jason Classon. Originally conceived as a multi-user chatroom, it quickly evolved into a photo-sharing platform. Only one year after Flickr.com had seen its release to a wider (web)audience, it was acquired by Yahoo!. After that Flickr expanded in various directions. New features were added in rapid pace and Flickr's user base grew exponentially. Through the increase in features and users Flickr became what it is today. At the moment of writing the Flickr platform enables users to:

- Upload photos and high-definition videos
- Organize photos with albums, keywords (tags) and description
- Crop, fix and edit photos online
- To put photos on the map (automatically by using a GPS-enabled device or manually by drag-and-dropping photos onto a map in Flickr's online interface)

When uploading photographs, a user also enters metadata like a title, tags, description, a geographic reference and comments. The rich metadata and Flickr's social network component made the photo-sharing platform evolve into an interface that enables 'exploring' the content. This can be done by clicking on a map, flicking through photos marked 'interesting' by others, or simply by searching or browsing specific keywords or themes.

It is clear that Flickr is first and foremost a platform of photo-enthusiasts. Discussions and comments on Flickr tend to focus on aesthetics and the actual depiction of the photo. This distinction is further enforced by Flickr's account policy. With a free account, a user can upload only 2 videos and 100MB of photos per calendar month. Users that are hitting these limits have the choice to upgrade to a paid account.

---

[7] http://www.telecompaper.com/news/article.aspx?cid=724530

Nevertheless, Flickr grew to a community of more than 40 million members that have uploaded more than 4 billion images altogether since its inception, often of high quality. A whole ecosystem of websites and applications has emerged that use and re-use photos sourced from Flickr. To support this, Flickr has released the first version of their application programming interface (API) already in August, 2004.[8]

Data Mining

Using the Ruby script language[9], a program is written that can download information through Flickr's API. We download metadata -- and not the actual photo -- of all geotagged photos of the city of Amsterdam (i.e. no sampling is done). This resulted in a database providing information on location, keywords, description, title and a time stamp (see Table 1) of uploaded photos. The geolocation of each photo allows us to look at the first dimension of place: *location*. The description, title and keywords tell something about *meaning*.

In addition, each photo contains a reference to the author or photographer of that photo. For each photographer, all available details (location, description, interests) of the photographer were added to the database as well (Table 2). We can use the location of the photographer to distinguish between local and tourist users and thus analyze the *heterogeneity* of place.

Data processing

To analyze the dimensions of place systematically, a number of quantitative GIS techniques are applied to the compiled database. A first step is to geocode the location of a user -- which means to assign also the user a geographic location on a map. In the metadata the user's location is a user-defined text field. Since filling out this field is optional, a user can enter anything. Sometimes people fill out 'Moon', 'Somewhere on earth' or other container terms which makes it difficult, and even impossible, to give the user a proper location on a map. On the contrary, real place names can be converted easily to a pair of coordinates using a geocoding service; the text input is compared with a geo-spatial database containing thousands of locations all around the world and, if a match exists, geographic coordinates are returned.

Secondly, to analyze patterns of density, we make use of a point density algorithm that is available in most GIS software. The input for this algorithm is a point map with all photo locations. The output is a grid with a cell size of 10 m. For each cell in this raster the number of points within that cell and its immediate surroundings (radius of 100 meters) is counted. This method is preferred to a crude count because the geographic location of each photograph is not really precise. For example, normal consumer GPS devices (like a

[8] An API is an interface that is implemented in a specific software (in this case Flickr's web platform) that enables other software to interact with that specific software program. For example, Flickr's API enables photo-management software to directly upload images to Flickr - without going through the Flickr.com website.

[9] http://www.ruby-lang.org/en/

smartphone) are inaccurate to several meters — especially in an urban context. In addition, user errors might be present as well (i.e. imprecise pinpointing on the map). By applying a wider search radius (e.g. 100 meters) the uncertainty of the photo-location can be taken into account.

The classification that is used in the visualization of the density maps is based on Jenks' natural breaks algorithm. It tries to minimize the within-class sum of squared differences, thus minimizing variance within each class. Although this leads to 'uneven' breaks, we argue that it is especially suitable to visualize distinct patterns of spatial clustering. Knowledge of small-scale clustering helps to get an understanding of which locations in the city are prominent in the dataset.

Thirdly, to quantify the uneven geography of uploaded photographs, we will look at the distribution of points. Therefore we will use a grid with a grid cell size of 600 by 600 meters. Counting all the points in each grid cell allows analysis of the uneven geographic distribution of photographs. Furthermore, to determine whether the point pattern could be a consequence of randomness, we look at spatial autocorrelation by using Moran's I — a ratio of 1 signals complete positive spatial autocorrelation and -1 complete negative spatial autocorrelation (cf. Smith et al. 2007 for more information).

Analysis of (perception of) place

We now turn to the dimensions of place in Amsterdam ingrained in volunteered geographic information; these are derived through data mining and processing of the metadata of Flickr photographs as described in the previous section. Before going into the details of place, we will first provide a profile of the users contributing to the rich source of information.

Who is photographing?

In total, 9,283 photographers have contributed a little more than 200,000 pictures in the city of Amsterdam. While on average a photographer uploads around 23 pictures, the median of 4 and mode of 1 in Table 3 and Figure 1 show that, in practice, the number of photos taken by each photographer does not follow a normal distribution. It looks more like a power-law or Pareto distribution. If we were to plot a power-law distribution on a log-log scale, it would follow a straight line. In our case, the distribution follows an almost straight line for most part of the graph (Figure 2), confirming that a minority of users contribute a majority of the pictures. For example, the user with most contributions uploaded more photographs than the 2419 least active users together.

Although it seems counterintuitive to take the trouble to create a Flickr account, geocode photos and then upload only a single photo in the Amsterdam area, there could be several good reasons for doing so. First of all, some users might upload hundreds of pictures to Flickr but just geocode one picture post-hoc to see ´how it works'. Second – and very likely, people use Flickr in different ways. Some users use it as a photo log or archive -- they upload every picture they take

-– and some are, like the top-user, 'genuine shutterbugs'. These are the few users that have contributed thousands of pictures. Others are more selective and might only create albums of their most precious pictures. It could very well be that only one or two of a user's most valued pictures were shot in Amsterdam.

The analysis of personal characteristics of the user is restricted to the information that has been provided by the user in the profile page. There is no way of telling gender, income or any other socio-economic characteristics. In fact, user accounts are highly anonymous. The location field is the only piece of information that tells us a little more about the user. Using the geocoding process described earlier, we were able to pinpoint 55% of the users on the map with reasonable accuracy and use this information to get an understanding of geographical differences.

Flickr is now available in 8 languages: English, Chinese, German, Spanish, French, Korean, Italian and Portuguese. In the top 15 of user origin countries, countries in which the mentioned languages are spoken as the official language are well represented (Table 4). Most countries in this list are relatively rich. Africa is highly underrepresented – in fact, the first African country on the list is South Africa (10 photographers). Asia and South America are also hardly present with only China, and Brazil in the top-15.

Of the 1355 Dutch photographers who have uploaded a photo referring to Amsterdam, 549 come from Amsterdam. These are the users that are labeled as 'local' in the rest of this chapter. All other users with a known home location -- be that somewhere else in the Netherlands or elsewhere in the world -- are 'tourists'. Local users have contributed 37323 photographs (30%), while tourists have uploaded 86108 pictures (70%).

Although we distinguished six dimensions of place in the analytical framework, we will only elaborate those that are clearly portrayed by the VGI data for Amsterdam, being location, meaning and heterogeneity.

Location - City

The first and most 'simple' characteristic of place is that place has a location. In this section we will therefore show where and how often photos are taken in Amsterdam. Just as the distribution of uploads per photographers, photos in Amsterdam are not evenly distributed in space (cf. Table 5). It is this uneven distribution -- high intensity in some areas, no photos at all in others -- that shows where place and space are constructed and consumed within the city of Amsterdam. Looking at Figure 3, it becomes immediately clear that the 'island' structure that we see in many American cities (Poorthuis, 2010) is less prominent in Amsterdam. Amsterdam is a different city; it is concentric with a small centre in the middle and larger and larger concentric circles around it. This is the structure that is also reflected by the density map. There are two centers of gravity: one around Central Station and the *Dam*. The other centre is located at *Leidseplein*. Other than that, we see a uniform spread of pictures in a circle around the centre, which turns into a less pronounced density on the other side

of the outer ring formed. Furthermore, there are several areas that are hardly photographed.

The maps deliberately do not portray the area of Amsterdam that lies at the periphery (outside of the ring road) as there are hardly any pictures taken there. The contrast between Amsterdam inside and outside of the *A10* ring road is often also found in vernacular references by students in Amsterdam. If somebody is asked where he or she lives, this is often done by asking 'Do you live inside or outside the ring road?' The answer 'outside of the ring road', for example in Slotervaart, will often be followed with a meaningful 'a-ha'. Apparently, living outside of the ring road or in North of Amsterdam has a certain image attached to it.

Like the uneven distribution of photographs across boroughs, the distribution of photographs per grid cell (see Table 6) shows a skewed distribution as well. About 28% of the grid cells in Amsterdam contain no pictures at all, while the most densely photographed cell (*Dam* and Central Station) -- covering 0.1% of the total area of Amsterdam -- contains 1.4% of all pictures taken in Amsterdam.

The uneven distribution of photographs that we see within Flickr is not problematic in and of itself. It is part and parcel of the very nature of how place is portrayed and produced within Flickr. The fact that a cluster of pictures can be found at *Leidseplein* while there are almost no photos in the surrounding neighborhoods says something about how that place is being perceived and produced by Flickr users. It is these peaks and valleys of photograph density that tell something about the perception and production of urban place.

This metaphor of a landscape of peaks and valleys is often used in the context of Volunteered Geographic Information[10]. However, we doubt whether this is the most apt metaphor available. The areas that are not photographed are not represented on the map – they are also not part of how a city is perceived by Flickr users. It is thus rather deceiving to characterize these areas as valleys, as they are completely off the mental map. Even the 'terra incognita' metaphor does not adequately describe this concept. The un-photographed areas are not uncharted or unexplored terrain – as such a notion holds an implicit wish to change that state and explore (or exploit) the said blank spot. If we think of the urban as a universe, then the densely photographed areas can be thought of as filaments – full of sparkling clusters and *galaxies* – and the empty, never photographed, areas are *voids* – nothing more than empty space. Empty space also means that place cannot exist there. Place has meaning – and in Flickr's cyberspace these voids have zero meaning. They are not photographed and thus they do not exist – they have no description, no keywords, no meaning, no sense of place, nothing at all.

Location - Local

[10] See for example: http://urbantick.blogspot.com/2010/01/new-city-landscapes-interactive.html

The clusters (of photographs) that emerge in a city-level analysis can also be seen at a micro-scale. The *Plantagebuurt*, for instance is home to the oldest zoo of the Netherlands, *Artis*. It is easy to spot it in the northeast corner of Figure 4. It is popular with both locals and tourists and receives over 1 million visitors per year[11]. But *Artis* is not the only 'hotspot' in the neighborhood: the *Wertheimpark* and the streets and bridges lining the *Amstel* river are also densely photographed. In the middle of the map is the *Roeterseiland* campus of the University of Amsterdam. The campus and its immediate surrounding seem to be a quiet oasis. The campus area is literally a stone's throw away from both *Artis* and the beautiful bridges over the *Amstel*, but somehow fails to capture the interest of Flickr users.

Meaning - City

In this section we will use the description of the actual photos to approximate the meaning of place. Table 7 shows the top 10 of most used keywords in Amsterdam. Not surprisingly, most keywords refer to geographic connotations (e.g. 'Amsterdam') although we also see keywords referring to time (2008, 2009). Further down the list, keywords become more descriptive. Architecture, clubbing, fun, and gay are all words that are used thousands of time to annotate photos in Amsterdam. Surprisingly, there are hardly any Dutch words among the keywords. The first Dutch word is 'uitgaan' [going out] on rank position 109. To capture the meaning of place, a number of meaningful categories is distilled:

- Art (museum, art, artist, kunst but NOT 'artis')
- Music (music, live music, concert, *muziek*)
- Nightlife (club, nightlife, nightclub, bar)
- Food (food, restaurant, *eten*)

According to Flickr users, art seems to be everywhere in Amsterdam, although a few specific hot spots can be discerned (Figure 5). In sharp contrast to the proliferation of art spots, there are only two -- very profound – music hot spots. Judging from this map, the city has not much more to offer. However, these photos also discount other live music venues that offer classical music, jazz and opera. Again, this might be a consequence of the specific characteristics of Flickr's user base (Jansen, 2010).

Nightlife is found in a wider array of places in different corners of the city. Food is to be found all over the city with concentrations in the centre areas. Noticeable is also that there are some minor hot spots outside of the inner ring as well. Apparently people venture further outside of the centre for good food.

Heterogeneity – City

[11] http://www.at5.nl/artikelen/11950/record-aantal-bezoekers-voor-artis-in-2008

Figure 6 and Figure 7 show the same density as discussed in the previous paragraph but distinguishes between local users and tourist users, which highlights some very clear differences. In general, tourist hot spots are very scarce outside of the *Stadhouderskade* ring. Local users seem to venture much more widely throughout the city. Since Amsterdam is a concentric city, standard distance is a useful measure to analyze the difference between the two. The standard distance for all tourist pictures is 2.6 kilometers. Photos taken by local users have a standard distance of 4 kilometers.

Tourists mainly visit and photograph the typical tourist places like *Dam* Square, Canal Belt and *Leidseplein*. The 'local' map is much more diverse. Naturally, the center is also well reflected, but a variety of hot spots can be found further away as well. When we compare specific meaning of place, this image becomes even more profound. In Figure 8 and 9 we look at nightlife areas for tourists and locals. The main nightlife areas for tourists are restricted to the typical tourist squares like *Leidseplein* and *Rembrandtplein*. Local users frequent completely different places, as can be seen in the *Westerpark* area, the *Jordaan*, *Nieuwmarkt*, and former *clubs 11* and *Twstd*.

Potentials, challenges and limitations

By linking existing theory on place with the empirical study of Volunteered Geographic Information, we stress that this bridging is not only useful but also necessary to successfully complete a meaningful analysis of the data at hand. It is in the bridging of these two bodies of work that new insights can be found: theory on place provides a useful framework to study VGI and the study of VGI can provide new insights into how place is produced and consumed in society.

Clearly, the buzz around VGI is not only empty hype. Key here is both the depth and breadth of the data at hand – resulting in some of the maps and findings presented in this paper. However, even though VGI has great potential for studying place, there are several challenges to consider. We showed that photographs are not evenly distributed through geographic space. Popular and affluent areas receive a disproportionate amount of attention and there are several issues around the aggregation of the data. What scale does the photo relate to? If place cannot be aggregated and averaged, how can the heterogeneous nature of place be represented in a straightforward and understandable way in research? In addition, the technological knowledge and skills needed to analyze the data could be a barrier to usage, especially when it comes to really utilizing the richness and contextual information of the data at hand.

Next to the potential and challenges, there are also specific limitations. First, for the analysis of place, we were restricted to the usage of 'geocoded' information, which forms only a subset of the total available data (~4% of all Flickr photographs are geocoded). Second, due to issues of representation, we always have to be aware of: 'whose place are we looking at'? And third, as discussed earlier, place has fuzzy and flexible boundaries. It can be as small as a room and as big as a city or the whole world. On the contrary, most VGI and spatial data is

much more rigid: it consists of sharp points, lines and polygons and does not allow for the fuzzy and flexibility that is inherent in place.

Coming from the specific potentials, challenges and limitation for the study of place, we will now turn to the more general potential, challenges and limitations for the use of VGI in the social science. The potential of VGI is obvious and has been discussed at length before (cf. Goodchild, 2007; Dykes et al., 2008; Purves et al., 2008). Specifically, it offers a large and diverse (enormous) dataset that is easy, and often freely accessible, while also offering rich, contextual data. It offers a 'different kind' of data than more conventional data sources (eg. census). Furthermore, usage of VGI could have an empowering effect on society since it takes away the monopoly of data production from governments and might motivate people to contribute, cooperate and work together towards a (common) goal.

The challenges and limitations of VGI are only slowly being uncovered. Naturally, since we move from expert (e.g. government) to 'amateur' production, issues of control arise. As a researcher, one has little control concerning quality. What kind of data is collected? Who is collecting data? Whose data is collected? Who is included and who is excluded? Who will use the data and for what purposes? There are two big challenges researchers can work towards to overcome. First, how do we deal with over- and underrepresentation within VGI? Can we let go the idea of an unbiased sample without losing scientific rigor? Second, how do we make sense of the data? Doing a shallow, large-N study of the data is relatively easy. But how can we really make use of the richness and contextual information of VGI to study social issues? To overcome these challenges, social scientists -- in cooperation with other disciplines -- need to come up with innovative methods to understand, analyze and make effective use of VGI.

Conclusions

VGI data proved to be a useful data source for studying production and consumption of place; the theoretical conceptualization and the utilization of VGI data enrich each other. Nevertheless, VGI data also puts up a number of challenges and limitations that should not be ignored when studying place and other issues. While VGI data opens up new opportunities for social science research, in particular with respect to the spatial perspective, we need to get a better understanding of how to overcome the imposed challenges and how to deal with the limitations like data quality and control.

Bibliography

Brenner, N., 2008. A thousand leaves: notes on the geographies of uneven spatial development. *Leviathan Undone*.

Cairncross, F., 1997. *The death of distance: how the communications revolution will change our lives*, Harvard Business Press.

Castells, M., 1997. The Power of Identity. The Information Age: Economy. *Society and Culture*, 2.

Cresswell, T., 2004. *Place: a short introduction*, Wiley-Blackwell.

Dodge, M., Kitchin, R. & Corporation, E., 2001. *Mapping cyberspace*, Routledge New York.

Dykes, J. et al., 2008. Exploring volunteered geographic information to describe place: visualization of the "Geograph British Isles" collection. *Proceedings of GIS Research UK*, p.256–267.

Elwood, S., 2010. Geographic information science: emerging research on the societal implications of the geospatial web. *Progress in Human Geography*, 34(3), p.349.

Friedman, T.L., 2007. *The world is flat: A brief history of the twenty-first century*, Picador USA.

Fukuyama, F., 1992. The end of history. *New York*.

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

Graham, S., 2002. Bridging urban digital divides? Urban polarisation and information and communications technologies (ICTs). *Urban Studies*, 39(1), p.33.

Harvey, D., 1996. *Justice, nature and the geography of difference*, Wiley-Blackwell.

Hugill, P.J., 1999. *Global communications since 1844: Geopolitics and technology*, Johns Hopkins Univ Pr.

Jansen, D., 2010. *Sociale media : Wat werkt voor wie op Flickr*. Master's Thesis. Universiteit van Amsterdam. Available at: http://dare.uva.nl/en/scriptie/341988.

Kitchin, Rob & Dodge, Martin, 2011. *Code/Space: Software and Everyday Life*, The MIT Press.

Massey, D.B., 1994. *Space, place, and gender*, Univ of Minnesota Pr.

Mitchell, W.J., 1995. *City of bits*, Mit Press Cambridge, MA.

Perkins, R. & Neumayer, E., 2009. Is the internet really new after all? The determinants of telecommunications diffusion in historical perspective. *The Professional Geographer*.

Purves, R.S. & Edwardes, A.J., 2008. Exploiting Volunteered Geographic Information to describe Place. In *GISRUK 2008 the 15th national GIS research conference for the UK, Manchester*.

Sack, R.D., 1997. *Homo geographicus*, Johns Hopkins University Press.

Scott, J.C., 1998. *Seeing like a state*, Yale University Press New Haven CT.

Smith, M.J. de, Goodchild, Michael F & Longley, P.A., 2007. *Geospatial Analysis: A Comprehensive Guide to Principles, Techniques and Software Tools* 2nd ed., Troubador Publishing Ltd.

Sui, D., 2008. The wikification of GIS and its consequences: Or Angelina Jolie's new tattoo and the future of GIS. *Computers Environment and Urban Systems*, 32(1), p.1–5.

Symanski, R., 2002. *Geography inside out*, Syracuse Univ Pr.

Taylor, P.J., 1999. Places, spaces and Macy's: place-space tensions in the political geography of modernities. *Progress in Human Geography*, 23(1), p.7.

Tuan, Y.F., 1977. *Space and place*, University of Minnesota Press.

Warf, B., 2001. Segueways into cyberspace: multiple geographies of the digital divide. *Environment and Planning B*, 28(1), p.3–20.

Zook, M. et al., 2004. New digital geographies: information, communication, and place. *Geography and technology*, p.155–176.

Tables

Field	Description
accuracy	accuracy of geocoding
datetaken	date that photo is taken (if unknown datetaken = dateupload)
datetaken granularity	precision of datetaken (i.e. day, month, year)
dateupload	date of photo upload
id	photo id
latitude	latitude of photo location
longitude	longitude of photo location
owner	owner id
ownername	owner name
tags	tags (can hold multiple tags)
title	title
description	description

Table 1: Metadata fields

Field	Value
ownerid	68364725@N00
ownername	JohnDoe
firstdatetaken	2006-03-07 11:24:58
count	41
location	“Alkmaar”

Table 2: Sample owner record

Measure	Value
Number of photos per photographer	
mean	23.05
median	4
mode	1 (N=2419)
standard dev	93.6
maximum	3290
Total	
number of photos	213990
number of photographers	9283

Table 3: Descriptive statistics photos per photographer

Country	Number of Photographers	Percentage of total
Netherlands	1355	27.29
United States	807	16.25
United Kingdom	545	10.97
Italy	323	6.50
Spain	316	6.36
Germany	226	4.55
France	190	3.83
Canada	147	2.96
Belgium	102	2.05
Brazil	88	1.77
Switzerland	65	1.31
Australia	58	1.17
China	52	1.05
Ireland	46	0.93
Finland	45	0.91

Table 4: Top 15 of origin countries

Moran's I	0.67
Z Score	27.2
P-value	0.000000

Table 5: Moran's I

Statistic	Number of photos per cell
Median	5
Mean	288
Standard Deviation	1730
Max	30357

Table 6: Descriptive statistics on number of photos/photographers per cell

Keyword	Number of mentions
amsterdam	148744
netherlands	45575
holland	36273
europe	14168
nederland	12490
2008	10437
2009	10136
geotagged	9030
thenetherlands	7824
canal	7137

Table 7: Top-10 of most mentioned keywords for Amsterdam

Figures

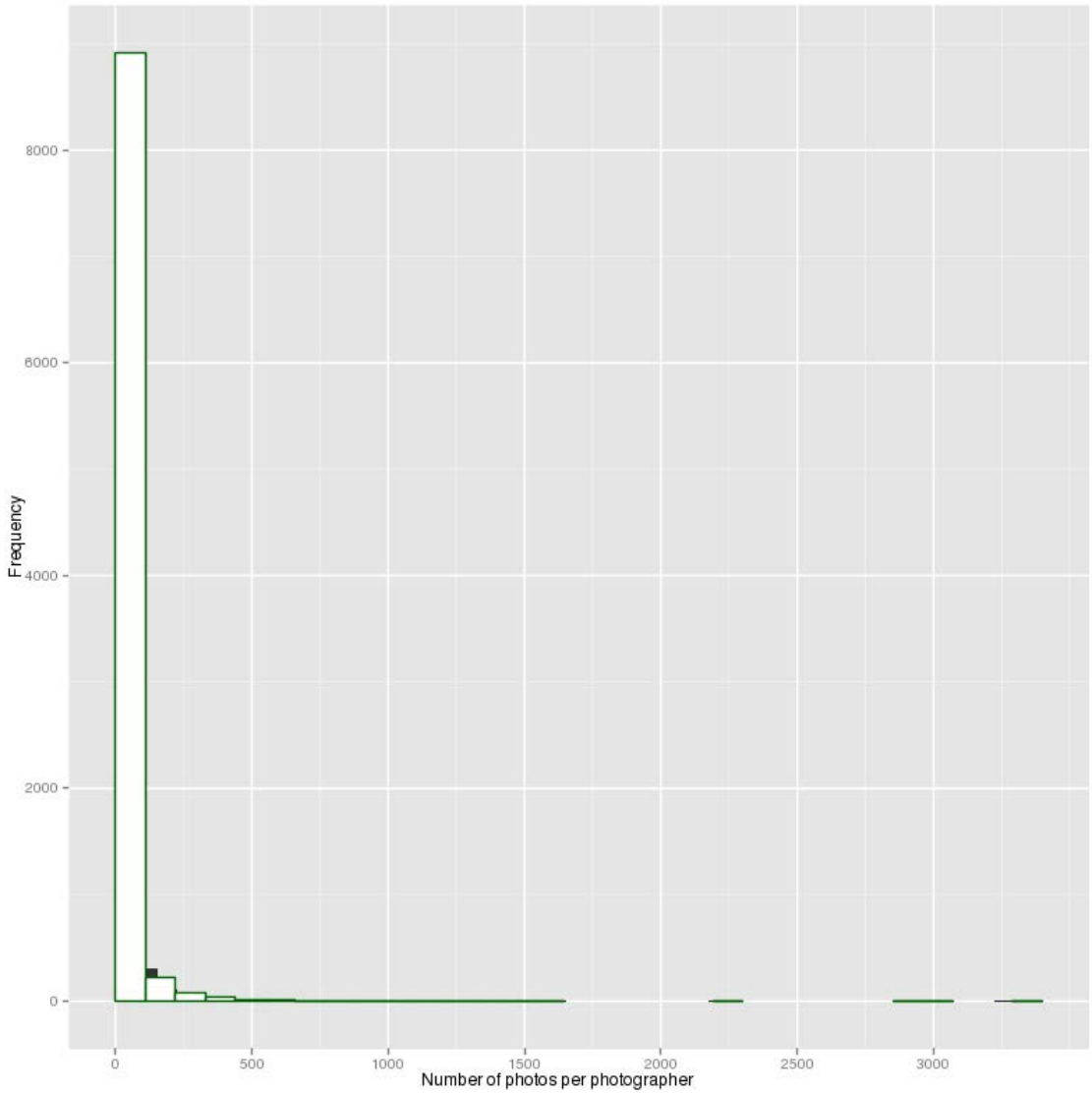


Figure 1: Number of photos per photographer

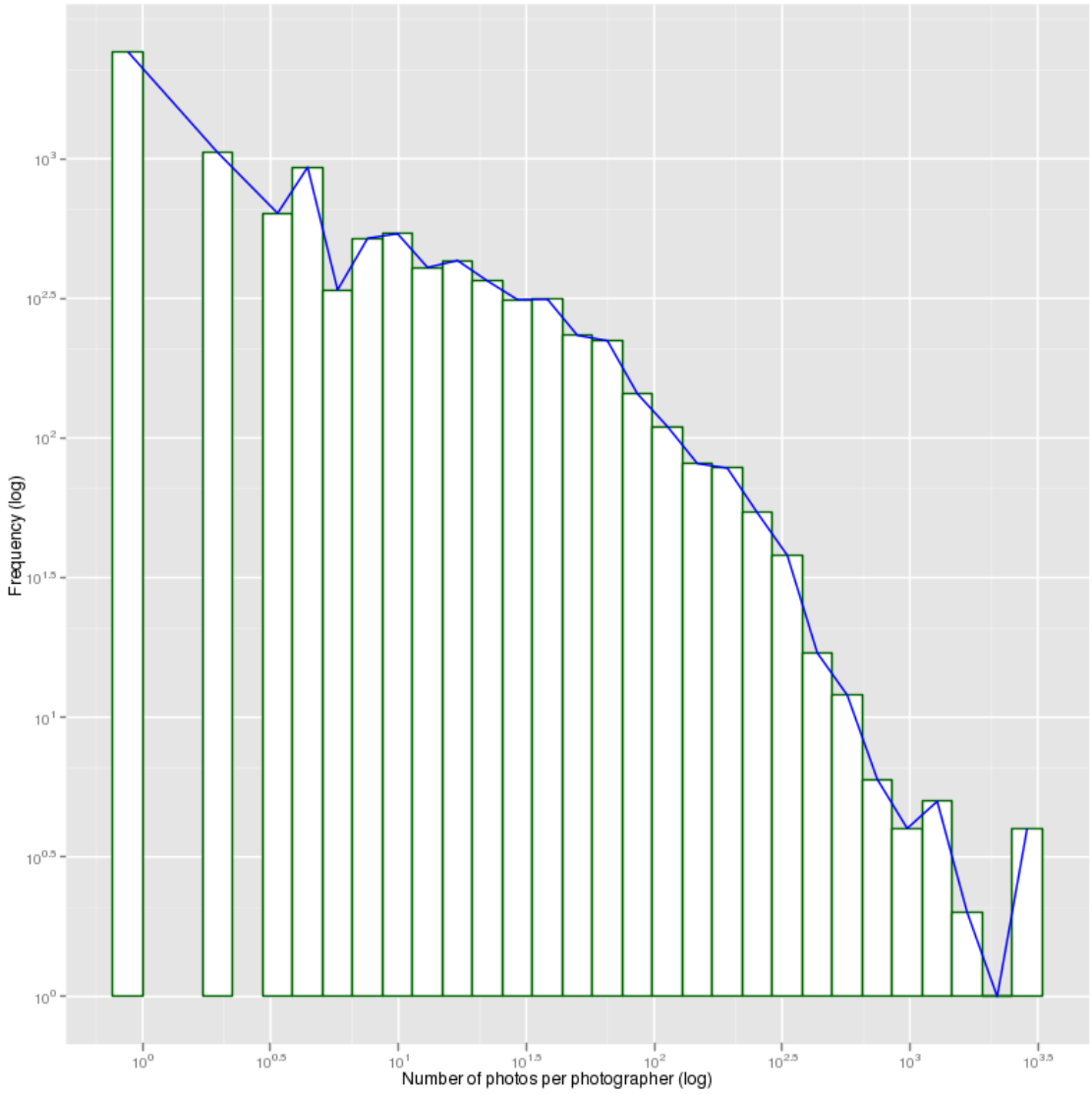


Figure 2: Number of photos per photographer (log/log)

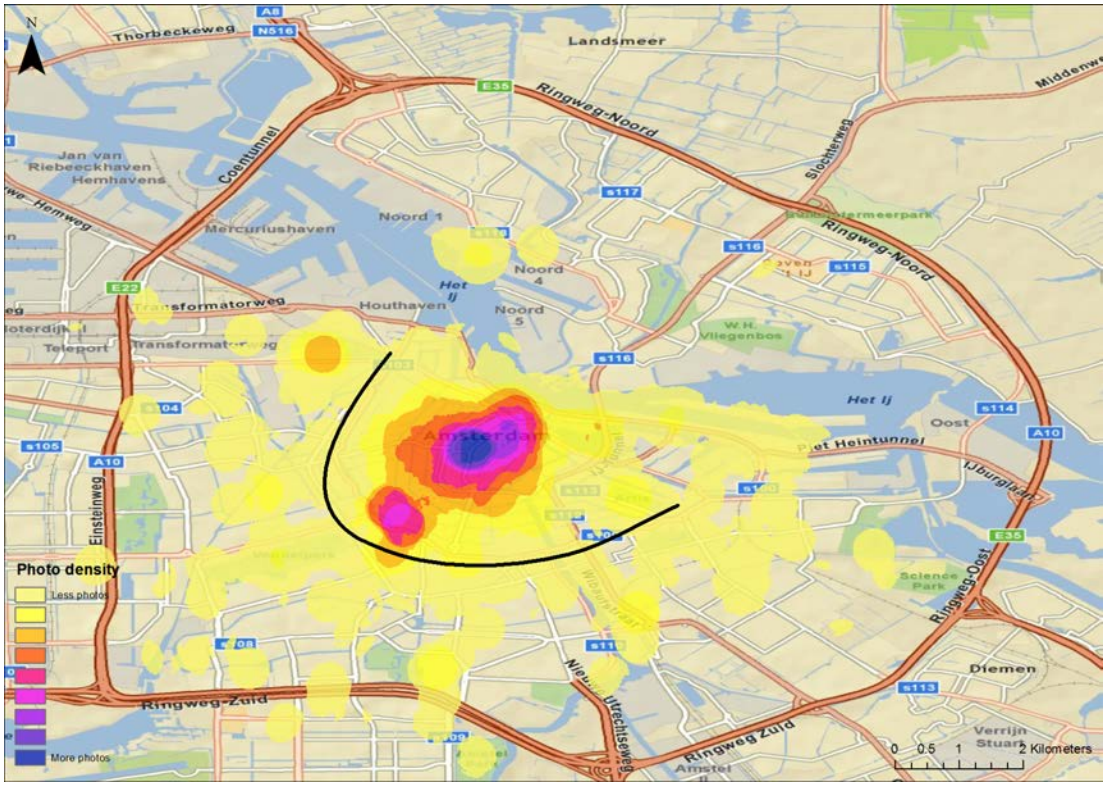
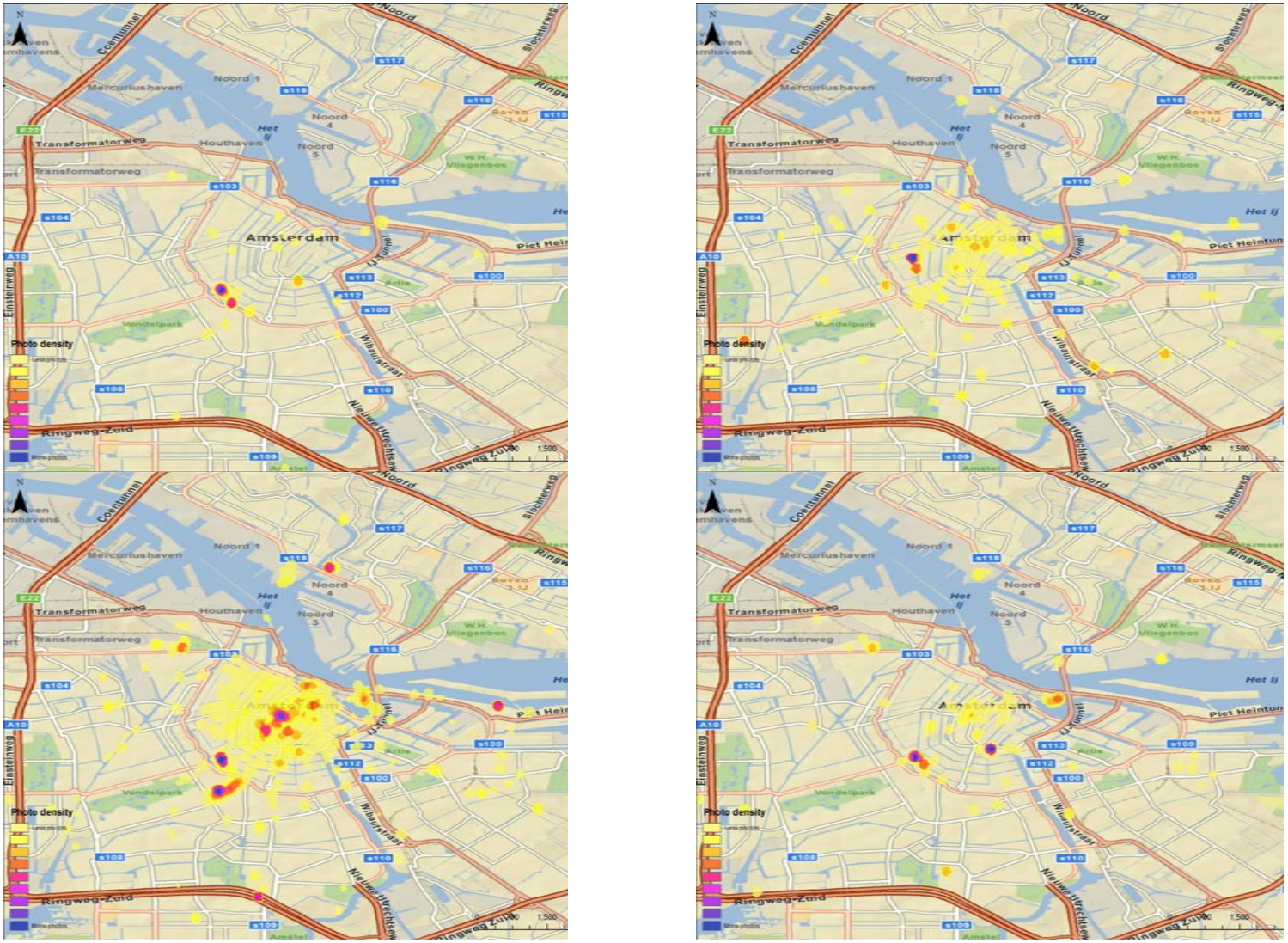


Figure 3: Density plot of photos taken in Amsterdam. Black line is *Nassaukade* ring.

Figure 4: *Plantagebuurt*. Every dot represents one photo

Figure 5: Clockwise: art, music, nightlife,



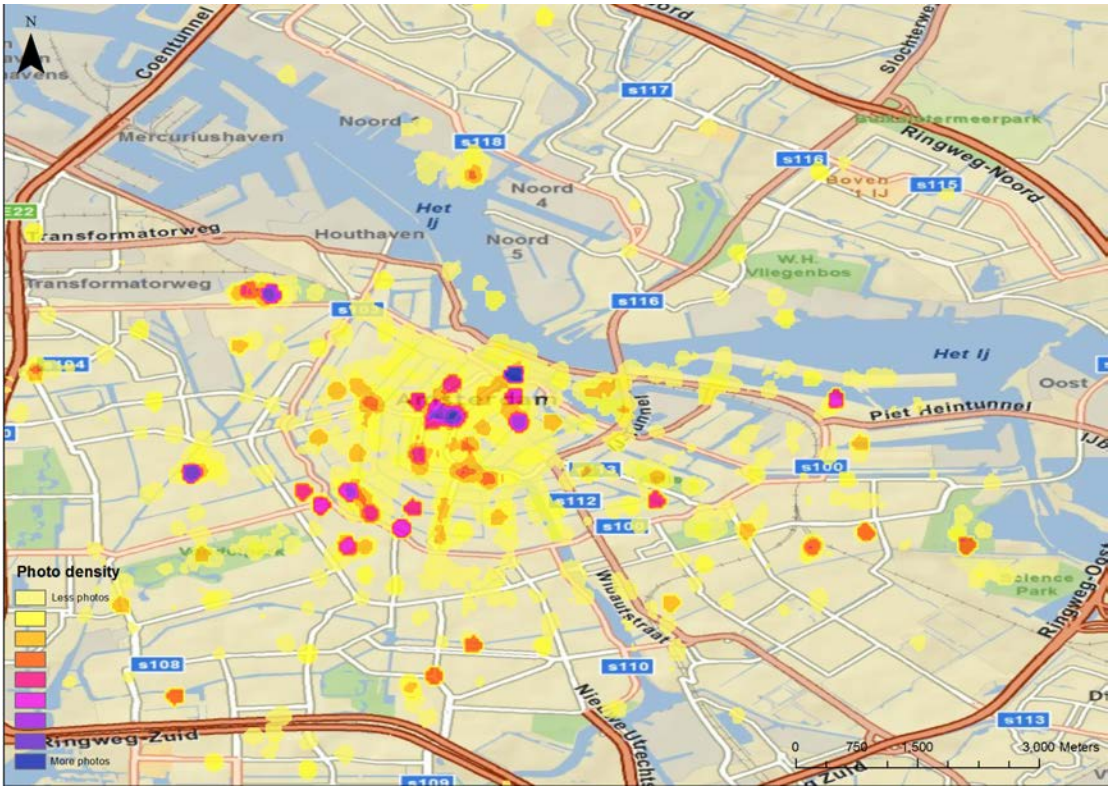


Figure 6: Photograph density - locals

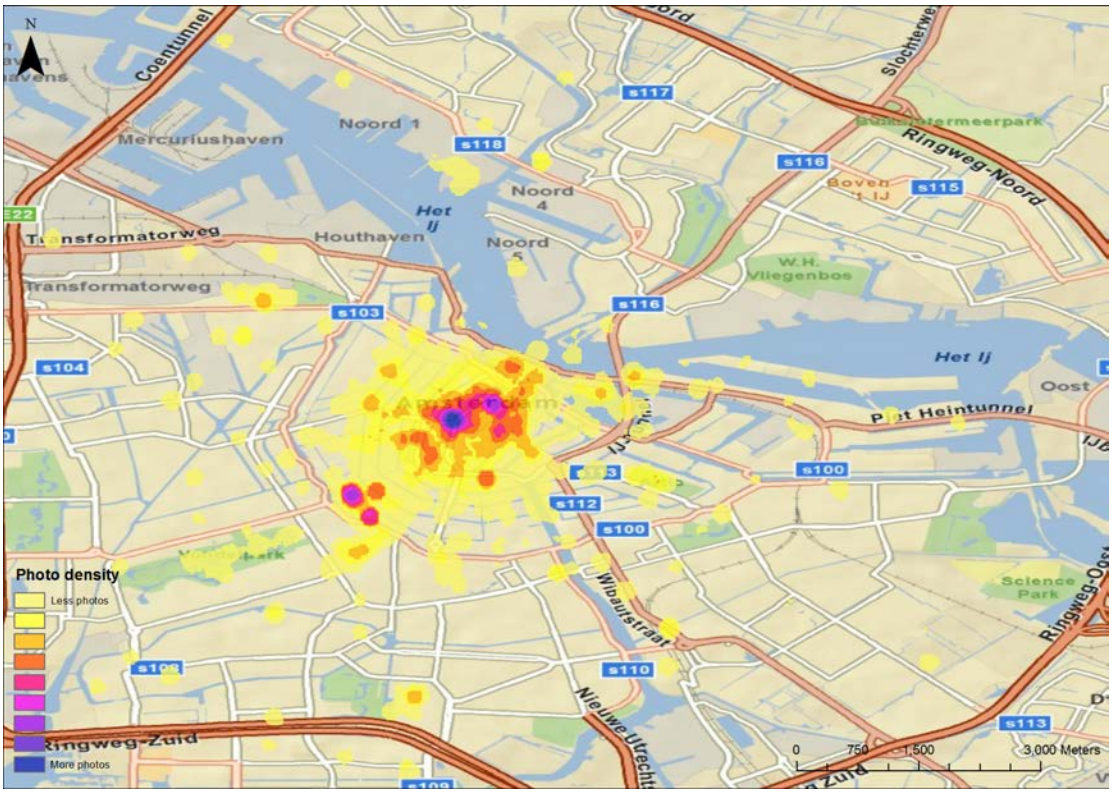


Figure 7: Photograph density – tourists

Figure 8: Nightlife density - locals

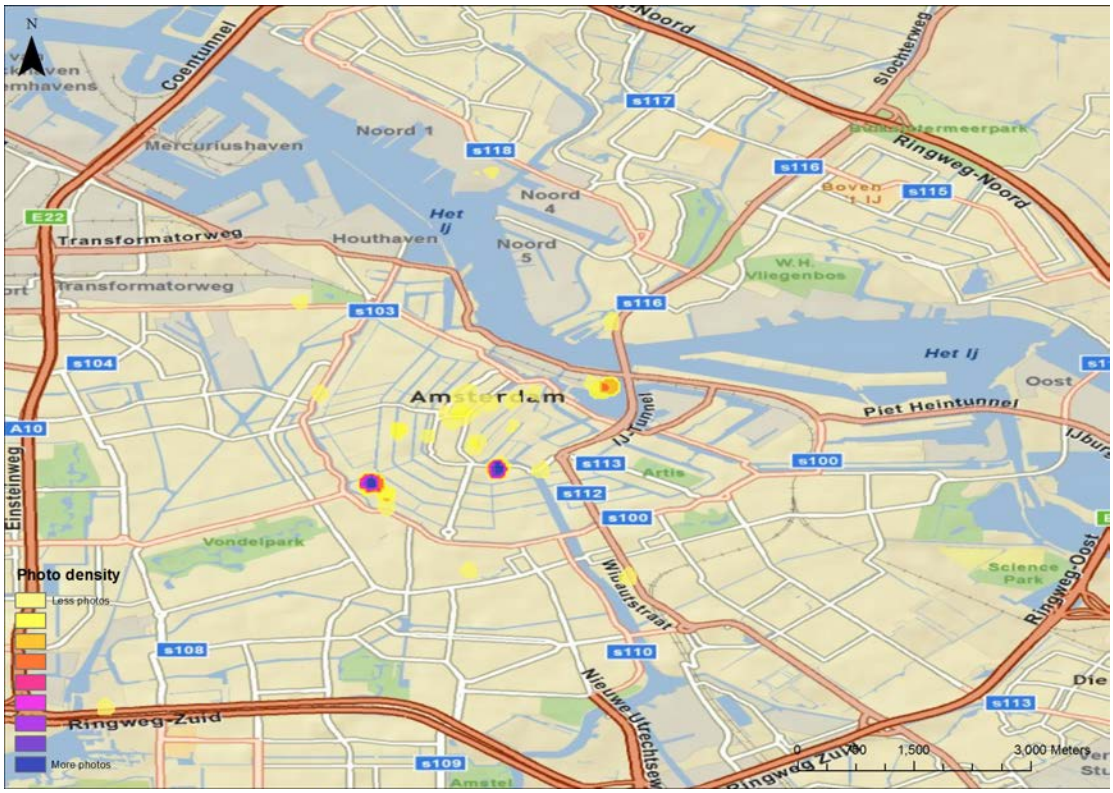


Figure 9: Nightlife density - tourists