
This is an electronic reprint of the original article.
This reprint may differ from the original in pagination and typographic detail.

Nummi, Pilvi

Social media data analysis in urban e-planning

Published in:
International Journal of E-Planning Research

DOI:
[10.4018/IJEPR.2017100102](https://doi.org/10.4018/IJEPR.2017100102)

Published: 01/10/2017

Document Version
Publisher's PDF, also known as Version of record

Please cite the original version:
Nummi, P. (2017). Social media data analysis in urban e-planning. *International Journal of E-Planning Research*, 6(4), 18-31. <https://doi.org/10.4018/IJEPR.2017100102>

This material is protected by copyright and other intellectual property rights, and duplication or sale of all or part of any of the repository collections is not permitted, except that material may be duplicated by you for your research use or educational purposes in electronic or print form. You must obtain permission for any other use. Electronic or print copies may not be offered, whether for sale or otherwise to anyone who is not an authorised user.

Social Media Data Analysis in Urban e-Planning

Pilvi Nummi, Aalto University, School of Engineering, Department of Built Environment, Espoo, Finland

ABSTRACT

Computational social media data analysis (SMDA) is opening up new possibilities for participatory urban planning. The aim of this study is to analyse what kind of computational methods can be used to analyse social media data to inform urban planning. A descriptive literature review of recent case study articles reveal that in this context SMDA has been applied mainly to location based social media data, such as geo-tagged Tweets, photographs and check-in data. There were only a few studies concerning the use of non-place-based data. Based on this review SMDA can provide planners with local knowledge about people's opinions, experiences, feelings, behaviour, and about the city structure. However, integration of this knowledge in planning and decision-making has not been completely successful in any of the cases. By way of a conclusion, a planning-led categorization of the SMDA method's tools and analysis results is suggested.

KEYWORDS

Citizen Participation, Computational Methods, Content Analysis, Literature Review, Location Based Social Media, Machine Learning, Network Analysis, Spatial Data Mining, Text Mining, Urban Planning

INTRODUCTION

The growth of Internet and social media use is providing new opportunities for communicating with and understanding local communities and people. Increasing mobile use and location based social media services are providing a huge social data source that contains data about people's behaviour, mobility and feelings about places. Social media data analysis methods have been widely developed and studied in other research areas, such as political science and commercial fields like marketing. For example, some location based data analysis methods have been developed for understanding business related issues like tourist travel behaviour and mobility patterns (Zheng, Zha, & Chua, 2012). In recent years, interest in these methods has also been increasing in urban planning practice. For example, in Turku, Finland, a case study was carried out that aimed at providing information about user mobility and place based experiences (Cerrone, Pau, & Lehtovuori, 2015) with geo-tagged social media data.

In this article, recent academic social media case studies are reviewed and analysed in order to gain an overall picture of the use of SMDA in urban e-planning. This review will bridge the gap between urban planning and urban computing, which has its theoretical background in the computational sciences. Understanding the computational methods in detail is not the focus of this article. Instead, this study concentrates on the applicability of the data analysis methods and their end-results with respect to urban planning. In general, it is argued that data analysis does provide information that can be used in planning, but that there is an evident gap between SMDA and urban planning practices.

In this introduction, I will present the theoretical background for this study, participatory urban planning, and define the concept of social media in this study and related works. The rest of the article

DOI: 10.4018/IJEPR.2017100102

Copyright © 2017, IGI Global. Copying or distributing in print or electronic forms without written permission of IGI Global is prohibited.

is organized as follows. Section 2 presents the research method, a descriptive literature review and the limitations of the study. The results (section 3) comprise a summary of the reviewed articles, such as the data and methods used. Section 4 elaborates the results and presents the classification of the methods of analysis from the viewpoint of participatory urban planning practices. Finally, section 5 discusses the findings and section 6 concludes the article.

Contemporary Participatory Urban Planning

Urban planning is a multidisciplinary and communicative practice that aims to create and foster liveable, sustainable, functioning and successful cities and areas. Planning is inseparably linked to decision-making and public policy making. For example, Eräranta and Staffans (2015) point out that rapidly changing conditions challenge urban planning practices striving to achieve more effective planning processes. They point out that issues like urbanization, climate change, globalization and societal differentiation are challenges that increase the need for better insight into the city and its users and the need for a holistic understanding of the planning ecosystem. The need for better informed planning and decision-making has been emphasized by other scholars in the urban planning field, too (Campagna, 2014; Campagna, Floris, Massa, Girsheva, & Ivanov, 2015; Evans-Cowley & Griffin, 2011).

The focus of planning is in most cases to change, create or to make it possible to create a place or area that can better serve the needs of people – economically, culturally and socially. An urban planner must have insight into the current situation of the place and the desired change in that place. Participatory urban planning is a user-centred approach that requires knowledge about how people use and experience places, and why they do so (Eräranta & Staffans, 2015). Digitalization is providing new possibilities for acquiring better knowledge, based on various data sources. Planning and decision support systems are information management systems that enable all stakeholders to get access to planning information (Eräranta & Staffans, 2015). User generated social media data can be seen as a valuable data source for planning as well.

The participatory planning approach is connected to deliberative democracy and the so-called communicative turn in urban planning. Pluralistic public discussion and argumentation are means to legitimate problem-solving in the deliberative model (Bäcklund & Mäntysalo, 2010). Communicative planning theory emphasizes the role of different stakeholders in the practice underlining the process and the content of planning (Healey, 1999). Building on the Habermasian idea of communicative action, communicative theory aims to reach a consensus between stakeholders. The ideal of a consensus-oriented process has been strongly criticized by theorists, because in planning, conflicting viewpoints and values always exist. However, communicative planning theory “defines an important role for the citizens as actors contributing to planning argumentation” (Bäcklund & Mäntysalo, 2010, p. 341).

Participatory planning theorists argue that knowledge is no longer only a domain of experts (e.g. planners or scientists). Instead there are multiple ways of knowing or multiple rationalities that relate to a multiplicity of different actors (Healey, 1999; Rydin 2007). In the concept of collaborative rationality presented by Innes and Booher (2010), rationality is formulated in interaction with different stakeholders as a learning process that includes local knowledge. In contrast to expert or administrative knowledge, local knowledge is informal and produced by local actors. It consists, for example, of opinions and memories, but experiential place-based knowledge can also be included in that concept (Rantanen & Kahila, 2009).

Based on this literature review, it is evident that social media is a rich source of user generated data and computational data analysis tools can provide us with information about people’s actions and behaviour in the urban space, which can be used in the democratic decision-making process. However,

it still seems that the integration of these tools and the knowledge produced with the tools in urban planning and decision-making is complicated. Despite the celebration of multiple epistemologies in participatory planning theories, it is still argued that there is no space for local knowledge in planning processes (Rydin 2007). However, the question of integrating these tools in urban planning is not only a question of the tools. It also requires changes to planning processes and practices.

Social Media in Urban Planning

Social media is a broad concept and it is often understood in many different ways, there being a great number and variety of social media tools and platforms. Khan (2013) sums up the concept as follows: “an Internet-based technologies/tools/concept that allows the creation and exchange of user generated content while letting users establish (at least one of these) identity, conversations, connectivity (i.e. presence), relationships, reputations, groups, and share content” (p. 136). In this article, social media is understood broadly, covering all these possible uses, but in practice the studied articles mainly use the most popular tools, like microblogs (e.g. Twitter), content communities (e.g. image sharing services like Flickr, Instagram and Panoramio), and check-in services (e.g. Foursquare, Swarm and Gowalla).

In the urban planning context, social media can be considered both as an interaction tool and as a data source. Social media has been part of participatory practices for years and is well studied as a communication tool (Evans-Cowley & Griffin, 2011). As Campagna (2014) suggests, using social media for “listening” to the feelings and opinions of the local community and asking them about current planning proposals would be one possible use with respect to planning processes. Another common goal for the use of social media is to broaden participation. In addition, considering user generated social media data as a valuable data source and analysing that data with computational methods can provide an effective way to create an understanding of the public’s view on planning (Evans-Cowley & Griffin, 2011). These computational (i.e. automatic) methods are in focus in this article while noting that manual or computer aided analysis methods (such as content analysis with tools like NVIVO or Atlas.ti) are likewise applicable methods as long as the amount of data is moderate. Computational methods do have significant advantages when analysing large amounts of data. The computational approach to social media is also related to a part of urban computing, a research field that uses computational methods to extract meaningful insights into cities (Zheng et al., 2012).

As Evans-Cowley and Griffin (2011) argue, planning has not been at the forefront in applying new technologies from social media, but is rather adapting to the on-going change in communication networks. Therefore, urban planners have an excellent opportunity to learn from other disciplines and their experiments, as these methods may often be applicable to the urban planning context as well.

Current research reveals the lack of usable social media analysis tools and methods for urban planning. Campagna et al. (2015) argue that the lack of shared and reliable methods or an analytical framework is preventing the utilization of social media data. A concluding article (literature review) has been written on this field from the perspective of digital tool development. In that article, Tasse and Hong (2014) studied city planners as one possible user group of social media data, focusing on the possibilities of using social media geo-tagged data in planning practices. They believe that “geotagged social media data can offer city planners and developers better information that can be used to improve planning and quality of life in cities” (p. 67) by providing more detailed data about socioeconomic status, mapping the quality of life and mobility, and by extracting “design patterns” and comparisons of different city structures.

The use of social media, data analysis, and visualization methods are very topical issues in urban e-planning research. Data analysis methods for location based social media are widely presented in a book by Ciuccarelli, Lupi and Simeone (2014). In their study, the aim was to design an SMDA tool, a “sensing platform” easily accessible for urban planners and other users without specific technical skills. For example, Bingham-Hall and Tidey (2016) have studied how social media visualization can enhance local communities and help decision-making by providing insights on local issues. Visualization of social media data is closely related to SMDA, but is not the focus of this article.

This literature review complements the previous studies by providing a critical approach to computational SMDA from an urban planning perspective, and emphasizing the need for usable social media data analysis tools for planning. In this study, outcomes of SMDA are considered to be representations of local knowledge.

RESEARCH METHOD

This study is carried out as a narrative literature review, which is a form of descriptive literature review. The aim of such a review is to build an overall picture of the topic and to identify possible problems (Baumeister & Leary, 1997). In this study, the aim is to gather relevant empirical studies related to computational SMDA methods and urban planning in order to gain an understanding of the methods and tools used in the academic field. As the social media data analysis studies are carried out in different fields, such as computer science, geography or GIScience and urban planning, the aim of this study is also to bring together knowledge regardless of the field of study and evaluate it from the urban planning perspective.

The main the main goal is to focus on research articles describing empirical studies related both to urban planning and social media data analysis. The review covers those articles that are mentioned as being related to urban planning; to be precise, peer-reviewed scholarly papers with the keywords “social media” AND “urban planning”. From this search result, articles specifically related to automatic data analysis of social media data have been manually selected. Some additional articles have been added to the literature review based on the author’s previous knowledge and the feedback given on this manuscript. Additional social media data analysis studies can obviously be found with other keywords like urban studies or urban informatics.

The main research questions are: what kinds of automatic SMDA methods are used in academic research related to urban planning? And how can these methods be categorized to gain an overall picture of the methods used in studies related to urban planning? The data collected from the articles are: area of the study of the research, methods of analysis and data used, evaluation methods, findings and conclusions drawn based on the data analysis. A classification of urban planning related to automatic SMDA methods is created and described based on this knowledge. It stems from participatory urban planning and the concept of local knowledge produced by local people.

RESULTS

The amount of studied research articles related to urban planning and SMDA is 24. The majority (in total 14) of the articles are from the field of computer science and their connection with urban planning is quite weak. Of the other studies, six articles are from the field of geography and one (Sykora et al., 2015) was created by a multidisciplinary team, and are more closely related to urban planning. Only two (Dunkel, 2015; Evans-Cowley & Griffin, 2011) come from the field of urban planning. In addition, one article comes from the field of computational social science.

All the studied articles were published within the last five years, as the oldest ones are from 2011. The majority of the articles (12) are from 2015 – only two are from 2014, one is from 2013, two are from 2012, and are three from 2011. Four articles from 2016 have been included in this study, as the data retrieval was carried out in the spring of 2016. It is obvious that the amount of publications is increasing, as the topic is now timely.

Most of the studies use georeferenced (i.e. geo-tagged) data such as check-in data (e.g. Foursquare, Swarm, Gowalla) or geo-tagged photographs (e.g. Flickr) or Tweets. Only two studies were based on non-geographic data (e.g. Tweets or other microblogs) (Evans-Cowley & Griffin, 2011; López-Ornelas & Zaragoza, 2015). Computational methods used in these cases can be divided into two main approaches: content analysis methods, such as sentiment analysis and topic modelling, and

Table 1. Distribution of research area, data source and method of analysis used in the articles

Research areas	Data sources	Computational methods
Computer science: 14 Geography: 6 Urban Planning: 2 Computational social science: 1 Multidisciplinary: 1	Photographs (geo-tagged): 6 Twitter (geo-tagged): 6 Check-in data (geo-tagged): 8 Mixed or custom data: 3 Microblogs: 2	Content analysis methods: - Topic modelling: 4 - Sentiment analysis: 2 Spatial data mining methods: - Spatial analysis: 6 - Clustering: 10 - Spatio-temporal analysis: 4 Statistical methods: 4 Network analysis: 1

spatial data mining methods, such as clustering methods and spatial or spatio-temporal analysis. In addition to these, statistical methods and network analysis were applied in some cases. See Table 1.

In most of the studies, the aim was to develop a novel data analysis method or algorithm or to apply well studied methods to new data. Only a few articles presented a method of analysis applied in a planning process. For example, Evans-Cowley and Griffin (2011) analysed the contents of microblogs in relation to transportation planning in Austin, and Campagna and colleagues (Campagna, 2014; Campagna et al., 2015) applied a custom social media platform to collect and analyse user interests related to city planning in Cagliari, Italy.

Some kind of evaluation or validation of the method was presented only in 11 case studies. In general, computational validations, such as an evaluation of an algorithm with different datasets (Chen, Chiang, & Peng, 2016), or contrasting the results with data provided by urban planning (i.e. official land-use zones defined in plans) (Frias-Martinez & Frias-Martinez, 2014), or with real world data or ground truth (Le Falher, Gionis, & Mathioudakis, 2015; Steiger, Westerholt, Resch, & Zipf, 2015; Sun, Fan, Li, & Zipf, 2016; Wakamiya, Lee, & Sumiya, 2011), or comparisons with similar kinds of studies (Çelikten, Le Falher, & Mathioudakis, 2016a) are typical in computer science and geography. Cranshaw, Schwartz, Hong and Sadeh (2012) evaluated the results by interviewing local people and Evans-Cowley and Griffin (2011) carried out a survey made for decision makers. It is remarkable that in no case study was user or usability testing used (or reported) as an evaluation method.

Critical viewpoints regarding the data-analysis methods are presented in several articles. For example, Bocconi, Bozzon, Psyllidis, Bolivar and Houben (2015) argue that “social media data are not always sufficient for deriving a detailed and accurate view of activities within an urban area” (p.178) and Shelton, Poorthuis and Zook (2015) point out that social media data is not a replacement for other forms of social and spatial data, but in their case study it represents a potentially rich source for empirically-grounded counter narratives of the inequalities and popular socio-spatial imaginaries.

CLASSIFICATION OF THE ANALYSIS METHODS

Based on the results of the analysis and findings presented in the articles, SMDA methods were classified into five categories based on the end-results from a planning point of view. These categories are: Mining opinions, Extracting place-based experiences, Defining city structure, Understanding crowd behaviour, and Multi-purpose methods. The classification is not unambiguous, because one article or method can fit into several classes. The categories are described in the following sections.

Mining Opinions

Discovering people’s opinions and public discussions about on-going or future planning projects is a central task in urban planning. Based on two different case studies, opinions extracted from social media may help urban planners to understand people’s desires and also help them to make better

decisions (Evans-Cowley & Griffin, 2011; López-Ornelas & Zaragoza, 2015). These tools are also said to support the interactions between planning authorities and citizens (Evans-Cowley & Griffin, 2011).

Sentiment and user analysis were used in two studies in order to identify people's opinions about planning related issues. López-Ornelas and Zaragoza (2015) studied Tweets related to a topical construction project, the installation of a new airport in Mexico City. They emphasize the importance of encouraging different groups to participate, choosing the right tools and raising public awareness when using new technological tools for participation. López-Ornelas and Zaragoza (2015) also argue that identifying user groups by network analysis is important, as is knowing which group the opinions represent. By categorizing users based on their tweeting activity and relationships they were able to detect who the main actors were in the discussions related to the airport project. They point out that this kind of knowledge is important in urban planning because people are directly involved in the decision making.

Evans-Cowley and Griffin (2011) state that sentiment analysis was a functional method for extracting public opinions from microblogs in a transportation planning case in the City of Austin. The experiment was only partly successful; they were able to generate engagement with the public, but the method of analysis did not appear to be effective in supporting decision-making. The difficulty was in understanding the reasons behind the sentiments. The authors describe the problem as follows: "One of the key weaknesses of this project was the inability to translate the results of the engagement into stories that could resonate with decision-making" (p. 1).

Extracting Place-Based Experiences

Place-based experiences are related to people's personal experiences and feelings, such as a feeling of stress (Sykora et al., 2015), landscape values (Dunkel, 2015), urban soundscapes (Aiello, Schifanella, Quercia, & Aletta, 2016), and smells (Quercia, Aiello, & Schifanella, 2016). Dunkel (2015) extracted people's landscape perceptions from geo-tagged photographs from Flickr. He applied common spatial analysis tools to visualize sightlines on a map, evaluate perceptions and perceptual landscape values. The analysis not only reveals points of interest but also places from where the landscape is perceived. Dunkel argues that geo-tagged images provide "a supplemental source of information with an otherwise unavailable perspective on the perception of the environment by the general public" (p. 185).

In a case study in Toronto, Canada, perceived stress was studied based on geo-tagged Tweets. Sykora et al. (2015) argue that this method provides new possibilities for studying how people perceive stress in the built environment and to examine how the environment affects health and well-being. These kinds of studies can provide urban planning information about designs that would prevent stress-related illnesses. Both Sykora et al. (2015) and Dunkel (2015) argue that social media data may help to understand how people perceive their surroundings.

Quercia et al. (2016) point out that, traditionally, studies about sensory urban experiences have focused on the visual sense. Social media opens up new possibilities for extracting experiences related to other senses as well; urban smells and soundscapes have been studied through geotagged tweets and images (Aiello et al., 2016; Quercia et al., 2016).

Understanding Crowd Behaviour

Geo-tagged social media data provides a diverse source for understanding how people behave in the urban environment. Crowd behaviour has been analysed in several studies in order to extract information about mobility (Chen et al., 2016), points and areas of interest (Hu, Gao, Janowicz, Yu, Li, & Prasad, 2015), functional clusters (Kling & Pozdnoukhov, 2012), land use (Frias-Martinez & Frias-Martinez, 2014), and for tracking and predicting events (Ferrari, Rosi, Mamei, & Zambonelli, 2011), and learning about the characteristics of neighbourhoods in terms of activities and functionality (Le Falher et al., 2015).

Methods that aim to understand the activities, functionalities and characteristics of a city often focus on points of interests (POI). They are places that attract a number of actions – e.g. geo-tagged

venues (check-ins), photos or Tweets – and thus represent an interesting place, as agreed to by a number of users. For example, Hu et al. (2015) present a method for extracting the boundaries of areas of interest (AOI) from geo-tagged images. The areas of interest are calculated from points of interest with a clustering algorithm. They also provide a spatio-temporal method for analysing how the AOIs form and change over time. They argue that understanding the evolution of AOIs can help in understanding and developing the future of urban areas. Geo-tagged photographs were also used by Paldino, Bojic, Sobolevsky, Ratti and González (2015) to extract the attractiveness of cities. These methods can provide information for the service development of the cities by providing information about what has to be enhanced in cities and where it can be appropriate to increase services targeting different categories of users.

Transportation planning is an obvious field that can benefit from understanding crowd behaviour. Steiger et al. (2015) extracted people's whereabouts from geo-tagged Tweets. They were able to detect significant clusters indicating home- and work-related human social activities and mobility patterns. Chen et al. (2016) also present a method for tracking mobility. They prove that users with similar mobility evolution patterns can be grouped via a clustering process. Both argue that geo-tagged social media data can help to understand people's travel demands and thus support transportation planning to meet people's needs.

Le Falher et al. (2015) developed an algorithm to quantify the similarity of city neighbourhoods. The primary use case for their study was recommending locations in the city (e.g. what neighbourhood to live in or visit as a tourist) but they claim that the methods are also applicable in urban planning. They argue that by identifying neighbourhoods that are similar to each other, the method may help to understand citizens' activities and experiences.

Defining the City Structure

Different approaches aim to reveal the characteristics of cities or neighbourhoods based on people's activities in places. These methods could be categorized under the topic "Understanding crowd behaviour" as they rely strongly on the same data and the idea of urban dynamics. However, there is a significant difference: while crowd behaviour studies focus on people and their activities, city structure analyses focus on the city and the built environment. For example, Frias-Martinez and Frias-Martinez (2014) extracted realized land use based on geo-tagged Tweets. They were able to detect areas for business, housing, nightlife, leisure and weekend activities as well as industrial areas. When compared to planned land use, this analysis may provide new insights into how areas are actually used.

The interest in detecting similarities between different neighbourhoods is related to recommendation services that, for example, advise users on restaurants, museums and shops to visit. One example of this kind of application is the online rental service Airbnb. Scholars argue that this kind of approach can help planners to better understand phenomena like urban gentrification (Zheng et al., 2012). SMDA applications called Livehoods, Hoodsquare and GeoTopics are examples of tools that were used to shed light on neighbourhood detection analysis (Çelikten, Le Fahler, & Mathioudakis, 2016a, 2016b; Cranshaw et al., 2012; Zheng et al., 2012). All these scholars argue that this kind of analysis method may be valuable for urban planners trying to understand the city structure.

Wakamiya et al. (2011) present an early experiment in characterizing urban areas based on geo-tagged social media data. They analysed urban characteristics from crowd behaviour extracted from geo-tagged Tweets. The concept of "urban characteristics" is approached in this case from people's daily routines, such as commuting, working, eating, and drinking. Based on crowd activities, the characteristics of a place can be captured. For example, an area where people are mainly commuting and working would be estimated to be an office town. Wakamiya and colleagues state that crowd behaviour adds an important social aspect to urban characteristics and they present a model that combines three views – the physical appearance, landmarks in crowd minds, and crowd behaviour – into the overall urban characterization.

Also, Kling and Pozdnoukhov (2012) extracted urban structure and characteristics from geo-tagged Tweets and check-in data. They were able to analyse also time related aspects of crowd behaviour and produce spatio-temporal knowledge about activities in the studied places. Thus, this analysis can answer questions like what areas are similar in terms of activities at a given time period.

One significant urban problem, segregation (e.g. socio-spatial inequality), may also be extracted from social media data. Shelton, Poorthuis and Zook (2015) used geo-tagged Tweets to analyse the segregation of two neighbourhoods, the poor West End and the affluent East End in Louisville. The study showed that it was possible to identify residents' activities and mobility by analysing geo-tagged Tweets and to extract knowledge about intra-neighbourhood segregation, mobility and inequality from that data with the aid of computational methods. Also, they revealed a counterpoint to the conventional narrative that class and racial segregation in housing is mirrored by limited mobility. It was found that actually the wealthy East End side is more spatially segregated. Thus, the analysis shows that social media data is a potential data source for empirical studies related to different forms of inequality and may also provide counter arguments against conventional understandings of phenomena (Shelton et al., 2015).

Apart from the planning context, a corresponding study by Kafsi, Cramer, Thomee and Shamma (2015) was carried out to quantify the uniqueness of regions. They investigated a dataset of geotagged images in order to filter such keywords that specifically characterize the area from other content that may reflect surrounding areas or general themes. This method was not mentioned as being used in urban planning, but it could support a planner in trying to understand the image of a region or a city.

Multi-Purpose Methods

The complexity of the urban environment and planning necessitates the use of various methods and approaches to enhance the understanding of the planning context. Social Glass developed by Bocconi et al. (2015) represents a platform that supports analysis and visualization of heterogeneous urban data together with social media data by multiple methods and tools. By combining social media data retrieved from Twitter, Instagram and Foursquare with demographic information, it provides a tool to detect people's movement patterns, the popularity of places, and traffic issues as well as opinions and preferences related to specific venues. With these functions, Social Glass can be considered to be a multi-purpose tool for urban planning.

Also, Campagna et al. (2015) present a social media geographic information (SMGI) tool that offers a variety of methods for collecting and analysing geo-tagged data in order to understand different user centred aspects, such as opinions, perceptions, and behaviour. They argue that tools that combine different methods of analysis as well as various data sources can be used to express people's perceptions and opinions, and thus can add a pluralist perspective to planning and decision-making. They propose a tool that can be used in a GIS application commonly used in planning. Bringing social media content to the planner's desktop as a new data source may support the utilization of user-generated content in urban planning.

Limitation of the Methods

Many scholars (for example, Hu et al., 2015; Sun et al., 2016) admit the representativeness bias (e.g. exclusion or under- or overrepresentation of certain population groups) in social media data. It is unquestionable that social media users represent only a fraction of the entire population, and the situation may even vary over time as Hu et al. (2015) points out. Different social media data contain different biases: for example, in check-in data there is an age group bias (young people use check-in services) and a place category bias (people check in more often to commercial places than to residential places) (Sun et al., 2016), and on Twitter the bias is related to uncertainty surrounding the location and semantic information, because Twitter users are generally referring to past or future events rather than reporting an on-going situation (Steiger et al., 2015). As Sun et al. (2016) point out, problems with the use of social media data can also be related to low data density in some countries or areas.

The way people use social media varies as well. Steiger et al. (2015) argue that “people using Twitter have individual motivations to post information and are also perceiving, identifying and interpreting their surroundings differently from each other, deciding what is worthwhile tweeting” (p. 257), and thus they argue that Tweets may only be a weak indicator of the observations of the real-world.

Despite the evident biases, all of the articles included in this study report more or less promising results. For example, Sun et al. (2016) show that it is possible to analyse the city structure based on check-in data. Hu et al. (2015) and Dunkel (2015) agree that despite the biases, social media studies do reveal useful information based on a large number of users. Thus, social media data can be seen as an important supplemental source of information and in many cases it provides information that is unavailable or expensive to collect (Dunkel, 2015). Dunkel also emphasizes that social media can provide a more balanced assessment of the perceived landscape, and thus it may help to integrate public values into planning processes. Sykora et al. (2015) point out that “there is evidence to suggest that social media data produced by individuals in situ and in near real-time may provide novel insights into the nature and dynamics of individuals’ responses to their surroundings” (p. 80).

DISCUSSION

Only a minority of the case studies included in this review were related to actual planning processes. Campagna and colleagues (Campagna, 2014; Campagna et al., 2015) and Evans-Cowley and Griffin (2011) brought up the complexity of problem-solving and decision-making in planning processes. In general, the context of urban planning was not very widely discussed and understood in most of the articles. Thus, there is a need to increase discussion and research related to social media use and notably SMDA methods in urban planning.

Several articles proposed that geo-tagged social media data can provide urban planning with useful information about city structure and crowd behaviour. Often, these arguments were not supported with very strong reasons and it remained unclear how this argument was validated. Especially research carried out in the field of computer science often focused on the computational issues and no evaluation in relation to real life planning situations was carried out. Even if the argument was agreed to by several scholars, the usefulness of the tools and results should be evaluated by and with urban planners before such a conclusion is drawn.

In some cases, like in Le Falher et al. (2015), the initial goal of the research related more to business potentials (in tourism and housing branches) than urban planning. For example, analysing the similarities or characteristics of neighbourhoods may serve applications for selecting travel destinations or searching for future living environments. However, the same methods may also be applicable in the urban planning context, but further development and evaluation of these methods is needed to prove that.

When developing computational methods, there are always several choices and assumptions made throughout the development process. These assumptions have an effect on the end result. Lehman (1998) argues that there is always a gap between the real world and the computational representation. This gap is bridged by assumptions, such as the choice of algorithms, values for parameters and so on. These choices and assumptions may, for example, be related to the use context of the tool or the environment. For example, using Foursquare data to analyse the characteristics of a place relies very much on the venue classification in Foursquare. Despite the venue types having been crowdsourced with Foursquare users, it is not self-evident that this data “constitute a basic ingredient to characterizing neighbourhoods”, contrary to what Zheng et al. (2012) argue. In fact, a study by Kim, Ihm, and Myaeng (2014) shows that the predefined set of Foursquare categories has a “limited ability to represent the dynamics of the place” (p. 561). They developed a text analysis method to improve the categorization by capturing hidden semantics of places from the Foursquare venue data with successful results.

Another example of assumptions was described by Sun et al. (2016). They argue that knowledge about the cities studied is needed to be able to choose the right parameters for the analysis. In that

particular study, a clustering algorithm was used to detect and define the city centre or centres and an assumption about the amount of possible city centres had to be made during the analysis process. Thus, local knowledge about the city would have been needed. Also, Campagna et al. (2015) state that the data collection and analysis tools are very local in nature and an understanding of the local context is needed in the analysis process. This underscores also the need for multidisciplinary co-operation between computer scientists, geographers and urban planners.

It is obvious that in the urban planning context, the social media data extraction methods and results should be evaluated with urban planners and local people. Carrying out evaluations with local people, as Cranshaw et al. (2012) did in their study, is a valuable method for assessing the validity of the results. In that study, the analysis result (i.e. the Livehoods map) was presented to local people who were then asked to give feedback about the map. In addition to that, the usability and usefulness of the methods in terms of urban planning can be measured together with planners. In his work, Dunkel (2015) sought to provide planners with the necessary knowledge to personally explore, visualize and interpret this data. This is certainly the direction in which the development of these tools should aim in order to become better integrated in planning processes.

Although the majority of the case studies report promising results on using social media data analysis, critical viewpoints can also be found. Bocconi et al. (2015) argue that social media does not always provide accurate information about people's activities and Shelton et al. (2015) point out that social media should not be considered a replacement for other forms of data, but they still see it as a potentially rich source of data.

Social media data analysis fills an evident gap in the field of urban planning and urban studies. It provides new possibilities for extracting information about the social aspects of the city: how people behave, move and interact with each other. These kinds of studies were laborious and expensive to carry out before the age of social media. All the articles studied in this literature review prove that social media can provide information that is relevant for urban planning. Though there are important issues that should be considered when developing and using the methods.

Based on this study, it seems that scholars from other academic fields (e.g. computer science) are more likely to propose methods related to understanding the characteristics of the physical environment while urban planning academics are more interested in people and their needs and opinions at a more detailed level. It is natural that the planning field emphasises detailed information, because in planning and decision-making listening to people's opinions is a part of planner's daily work.

One important question in choosing the methods of analysis is whether we are interested in the city as a physical environment or in the citizens as users of the city. In planning, the physical place is self-evidently in focus, and therefore methods that provide end-results in map format are relatively easy to use in planning. However, other kinds of tools – like network analysis – may be better at providing information about people living in a particular place.

SMDA methods are opening new possibilities for incorporating user-generated content in urban planning. For example, extracting feelings from social media enables urban planners to create places that also feel good. It is argued that until now, mainly bad smells and unpleasant noises have been taken into consideration in planning. In the future, more fulfilling and attractive cities and environments could be created by using knowledge about good smellscapes and pleasant soundscapes. (Aiello et al., 2016; Quercia et al., 2016)

CONCLUSION

This literature review compiles and describes the current state of computational social media data analysis in urban planning. Based on this study, there is evidence that various SMDA methods can provide urban planning with relevant and new information about mobility, use of space, the characteristics of space, and people's behaviour and experiences. However, it is evident that the usability and usefulness of the analysis methods and tools should be further developed and evaluated

in real word planning situations. There is also a need to provide urban planning tools that can be used and developed by planners themselves (Dunkel, 2015).

Also, analysing the data-driven methods in terms of urban planning theories will be needed to gain a deeper understanding of the possibilities of social media in planning and participation. In participatory planning theory, knowledge building is seen as a part of the learning process in collaborative planning processes (Healey, 1999). From this perspective, SMDA can be seen as a tool for knowledge building and collaboration in urban planning.

A classification of the data analysis methods, based on their end results from an urban planning perspective, is proposed in this article. A classification helps us gain an overall understanding of the methods and tools. The suggested categories are:

- Mining opinions to find out people's opinions and desires as expressed in public discussions
- Extracting place-based experiences to understand urban experiences and feelings, such as landscape values, feeling of stress or experiences of sounds and smells
- Defining the city structure to understand the functionalities and characteristics of the built environment
- Understanding crowd behaviour, such as mobility and activities in urban environments
- Multi-purpose methods that combine different data sets and analysis tools to get more comprehensive insights into the complexity of cities

There are still many unrevealed possibilities in SMDA for urban planning. The majority of the methods studied in this review provide experiential or local knowledge about places in a certain timeframe, often in recent history. Information that is related to experiences, people's behaviour and the city structure is helpful in trying to understand the city or an area as it is in that moment of time. Still, planning is a future oriented action, and thus it is worth considering how SMDA could also support the effort to envision the possible futures of urban areas. For example, mining ideas or future visions from citizens' self-organized social media groups would provide planners with bottom-up oriented viewpoints that have not been steered by the administration. Also, advanced text analysis methods could provide a deeper understanding about opinions if the quality and strength of arguments could be extracted from data. While these are just examples of the possibilities that SMDA opens up with respect to urban planning, future research and further development of the methods and tools will probably reveal more applications that are beyond our imagination at the moment.

REFERENCES

- Aiello, L. M., Schifanella, R., Quercia, D., & Aletta, F. (2016). Chatty maps: Constructing sound maps of urban areas from social media data. *Royal Society Open Science*, 3(3), 1–19. doi:10.1098/rsos.150690 PMID:27069661
- Bäcklund, P., & Mäntyselkä, R. (2010). Agonism and institutional ambiguity: Ideas on democracy and the role of participation in the development of planning theory and practice – the case of Finland. *Planning Theory*, 9(4), 333–350. doi:10.1177/1473095210373684
- Baumeister, R. F., & Leary, M. R. (1997). Writing narrative literature reviews. *Review of General Psychology*, 1(3), 311–320. doi:10.1037/1089-2680.1.3.311
- Bingham-Hall, J., & Tidey, J. (2016). Visualizing social medias impact on local communities. *Visual Communication*, 15(3), 317–328. doi:10.1177/1470357216645710
- Bocconi, S., Bozzon, A., Psyllidis, A., Bolivar, C. T., & Houben, G.-J. (2015). Social glass: A platform for urban analytics and decision-making through heterogeneous social data. In A. Gangemi, S. Leonardi, & A. Panconesi (Eds.), *Proceedings of the 24th International Conference on World Wide Web (WWW '15 Companion)* (pp. 175–178). New York: ACM. doi:10.1145/2740908.2742826
- Campagna, M. (2014). The geographic turn in social media: Opportunities for spatial planning and geodesign. In B. Murgante, S. Misra, A. M. A. C. Rocha et al. (Eds.), *Computational science and its applications – ICCSA 2014* (pp. 598–610). Berlin: Springer. doi:10.1007/978-3-319-09129-7_43
- Campagna, M., Floris, R., Massa, P., Girsheva, A., & Ivanov, K. (2015). The role of social media geographic information (SMGI) in spatial planning. In S. Geertman, J. Ferreira, Jr., R. Goodspeed et al. (Eds.), *Planning support systems and smart cities: Lecture notes in geoinformation and cartography* (pp. 41–60). Basel: Springer International Publishing Switzerland. doi:10.1007/978-3-319-18368-8_3
- Çelikten, E., Le Falher, G., & Mathioudakis, M. (2016a). Modeling urban behavior by mining geotagged social data. *Transactions on Big Data*. DOI:10.1109/TBDDATA.2016.2628398
- Çelikten, E., Le Falher, G., & Mathioudakis, M. (2016b). “What is the city but the People?”: Exploring urban activity using social web traces. In J. Bourdeau, J.A. Hendler, R. Nkambou Nkambou et al. (Eds.), *Proceedings of the 25th International Conference Companion on World Wide Web (WWW '16 Companion)* (pp. 167–170). Geneva: International World Wide Web Conferences Steering Committee. doi:10.1145/2872518.2901922
- Cerrone, D., Pau, H., & Lehtovuori, P. (2015). *A sense of place: Exploring the potentials and possible uses of Location Based Social Network Data for urban and transportation planning in Turku City Centre*. Turku Urban Research Programme's Research Report 1/2015. Estonian Academy of Arts.
- Chen, C., Chiang, M., & Peng, W. (2016). Mining and clustering mobility evolution patterns from social media for urban informatics. *Knowledge and Information Systems*, 47(2), 381–403. doi:10.1007/s10115-015-0853-4
- Ciuccarelli, P., Lupi, G., & Simeone, L. (2014). *Visualizing the data city: Social media as a source of knowledge for urban planning and management*. Heidelberg: Springer. doi:10.1007/978-3-319-02195-9
- Cranshaw, J., Schwartz, R., Hong, J., & Sadeh, N. (2012). The Livehoods project: Utilizing social media to understand the dynamics of a city. *Proceedings of the International AAAI Conference on Web and Social Media* (pp. 58–65). Dublin: AAAI Press.
- Dunkel, A. (2015). Visualizing the perceived environment using crowdsourced photo geodata. *Landscape Architecture and Environmental Planning. Special Issue: Critical Approaches to Landscape Visualization*, 142, 173–186.
- Eräranta, S., & Staffans, A. (2015, July 7–10). From situation awareness to smart city planning and decision making. *Paper presented at the 14th International Conference on Computers in Urban Planning and Urban Management*, Cambridge, MA.
- Evans-Cowley, J. S., & Griffin, G. P. (2011). Micro-participation: The role of microblogging in planning. *Social Science Research Network*. Retrieved from <http://ssrn.com/abstract=1760522>

- Ferrari, L., Rosi, A., Mamei, M., & Zambonelli, F. (2011). Extracting urban patterns from location-based social networks. In C.S. Jensen, W. Lee, Y. Zheng et al., (Eds.), *Proceedings of the 3rd ACM SIGSPATIAL International Workshop on Location-Based Social Networks (LBSN '11)* (pp. 9–16). New York: ACM. doi:10.1145/2063212.2063226
- Frias-Martinez, V., & Frias-Martinez, E. (2014). Spectral clustering for sensing urban land use using Twitter activity. *Engineering Applications of Artificial Intelligence*, 35, 237–245. doi:10.1016/j.engappai.2014.06.019
- Healey, P. (1999). Institutional analysis, communicative planning, and shaping places. *Journal of Planning Education and Research*, 19(2), 111–121. doi:10.1177/0739456X9901900201
- Hu, Y., Gao, S., Janowicz, K., Yu, B., Li, W., & Prasad, S. (2015). Extracting and understanding urban areas of interest using geotagged photos. *Computers, Environment and Urban Systems*, 54, 240–254. doi:10.1016/j.compenurbysys.2015.09.001
- Innes, J., & Booher, D. (2010). *Planning with complexity: An introduction to collaborative rationality for public policy*. Oxon, United Kingdom: Routledge.
- Kafsi, M., Cramer, H., Thomee, B., & Shamma, D. A. (2015). Describing and understanding neighborhood characteristics through online social media. *Proceedings of the 24th International Conference on World Wide Web (WWW '15)* (pp. 549–559). Geneva: International World Wide Web Conferences Steering Committee. doi:10.1145/2736277.2741133
- Khan, G. F. (2013). The Government 2.0 utilization model and implementation scenarios. *Information Development 2015*, 31(2), 135–149.
- Kim, E., Ihm, H., & Myaeng, S. (2014). Topic-based place semantics discovered from microblogging text messages. In C. Chung, A. Broder, K. Shim et al., (Eds.), *Proceedings of the 23rd International Conference on World Wide Web (WWW '14 Companion)* (pp. 561–562). USA, New York: ACM. doi:10.1145/2567948.2576955
- Kling, F., & Pozdnoukhov, A. (2012). When a city tells a story: Urban topic analysis. In I. Cruz, C. Knoblock, P. Kröger et al., (Eds.), *Proceedings of the 20th International Conference on Advances in Geographic Information Systems (SIGSPATIAL '12)* (pp. 482–485). New York: ACM. doi:10.1145/2424321.2424395
- Le Falher, G., Gionis, A., & Mathioudakis, M. (2015). Where is the Soho of Rome? Measures and algorithms for finding similar neighborhoods in cities. *Proceedings of the International AAAI Conference on Web and Social Media* (pp. 228–237). England, Oxford: AAAI Press.
- Lehman, M. M. (1998). Software's Future: Managing Evolution. *IEEE Software*, 15(1), 40–44. doi:10.1109/MS.1998.646878
- López-Ornelas, E., & Zaragoza, N. M. (2015). Social media participation: A narrative way to help urban planners. In G. Meiselwitz (Ed.), *Social Computing and Social Media, LNCS* (Vol. 9182, pp. 48–54). Switzerland, Basel: Springer International Publishing Switzerland. doi:10.1007/978-3-319-20367-6_6
- Paldino, S., Bojic, I., Sobolevsky, S., Ratti, C., & González, M. C. (2015). Urban magnetism through the lens of geo-tagged photography. *EPJ Data Science*, 4(5), 1–17.
- Quercia, D., Aiello, L. M., & Rossano, S. (2016). The emotional and chromatic layers of urban smells. *Proceedings of the 10th International AAAI Conference on Web and Social Media (ICWSM)* (pp. 309–318). California, Palo Alto: AAAI Press.
- Rantanen, H., & Kahila, M. (2009). The SoftGIS approach to local knowledge. *Journal of Environmental Management*, 90(6), 1981–1990. doi:10.1016/j.jenvman.2007.08.025 PMID:18556113
- Rydin, Y. (2007). Re-examining the Role of Knowledge within Planning Theory. *Planning Theory*, 6(1), 52–68. doi:10.1177/1473095207075161
- Shelton, T., Poorthuis, A., & Zook, M. (2015). Social media and the city: Rethinking urban socio-spatial inequality using user-generated geographic information. *Landscape and Urban Planning*, 142, 198–211. doi:10.1016/j.landurbplan.2015.02.020
- Steiger, E., Westerholt, R., Resch, B., & Zipf, A. (2015). Twitter as an indicator for whereabouts of people? Correlating Twitter with UK census data. *Computers, Environment and Urban Systems*, 54, 255–265. doi:10.1016/j.compenurbysys.2015.09.007

Sun, Y., Fan, H., Li, M., & Zipf, A. (2016). Identifying city center using human travel flows generated from location-based social networking data. *Environment and Planning. B, Planning & Design*, 43(3), 480–498. doi:10.1177/0265813515617642

Sykora, M., Robertson, C., Shankardass, K., Feick, R., Shaughnessy, K., Coates, B., . . . Jackson, T. W. (2015). Stresscapes: Validating linkages between place and stress expression on social media. In I. Katakis, F. Schnitzler, T. Liebig et al., (Eds.), *Proceedings of the 2nd International Workshop on Mining Urban Data*, Lille, France (pp. 80–84).

Tasse, D., & Hong, J. I. (2014). Using social media data to understand cities. In P. Thakuriah, N. Tilahun, & M. Zellner, (Eds.), *Proceedings of NSF Workshop on Big Data and Urban Informatics* (pp. 64–79). Chicago: University of Illinois Press.

Wakamiya, S., Lee, R., & Sumiya, K. (2011). Crowd-based urban characterization: extracting crowd behavioral patterns in urban areas from Twitter. In C.S. Jensen, W. Lee, Y. Zheng et al., (Eds.), *Proceedings of the 3rd ACM SIGSPATIAL International Workshop on Location-Based Social Networks (LBSN '11)* (pp. 77–84). New York: ACM. doi:10.1145/2063212.2063225

Zheng, Y.-T., Zha, Z.-J., & Chua, T.-S. (2012). Mining Travel Patterns from Geotagged Photos. *ACM Transactions on Intelligent Systems and Technology*, 3(3), 1–18. doi:10.1145/2168752.2168770

Pilvi Nummi architect M Sc. is a doctoral student in Aalto University, School of Engineering, Department of Built Environment. She works also as a town planning architect in the municipality of Sipoo (Finland). Her research interest is in e-Planning and e-Participation in urban and land-use planning.