

# Integrating Authoritative and Volunteered Geographic Information for spatial planning

Pierangelo Massa\* and Michele Campagna

University of Cagliari, DICAAR, Via Marengo 2, Cagliari 09123, Italy,

\*pmassa@unica.it

## Abstract

This contribution concerns ongoing research by the authors on the integrated use of Social Media Geographic Information (SMGI) and Authoritative Geographic Information (A-GI) as a support in urban and regional planning. Advances in Information and Communication Technologies (ICT) are fostering the production and the sharing of georeferenced user-generated contents, namely Volunteered Geographic Information (VGI) and SMGI, which may complement traditional spatial data sources. VGI is a voluntary contribution by users in order to collect or to disseminate geographic knowledge, while SMGI may be considered a deviation from VGI nature, due to the implicit and passive mode in disseminating geographic information, which is exclusively one embedded attribute of the main shared information. However, SMGI may offer unprecedented opportunities to investigate users' needs, opinions, behaviors and movements, thus representing a potential support for analysis and decision-making in spatial planning. In this respect, the authors present an original tool called Spatext, which allows collection, management and analysis of SMGI in GIS environment, easing the integration of SMGI with official information. Afterwards, the opportunities for spatial planning arising from

---

### How to cite this book chapter:

Massa, P and Campagna, M. 2016. Integrating Authoritative and Volunteered Geographic Information for spatial planning. In: Capineri, C, Haklay, M, Huang, H, Antoniou, V, Kettunen, J, Ostermann, F and Purves, R. (eds.) *European Handbook of Crowdsourced Geographic Information*, Pp. 401–418. London: Ubiquity Press. DOI: <http://dx.doi.org/10.5334/bax.ac>. License: CC-BY 4.0.

SMGI are demonstrated through a case study where this type of information is used for investigating the geography of places. An original methodology is developed applying clustering techniques on the spatial and the temporal components of SMGI collected from Instagram. The applied methodology enables the identification of residential buildings that are not mapped in available official datasets. The results demonstrate how SMGI may be proficiently used to integrate and update A-GI, as well as to investigate the users' behaviors and movements in an urban environment.

### Keywords

SMGI, VGI, Urban planning, Spatial planning, Instagram

### Introduction

In recent years, continuous advances in Information and Communication Technologies (ICT), the internet and Web 2.0 technologies are strengthening the production, the sharing and the access of user-generated contents among millions of users worldwide. The availability of user-generated contents may represent a novel source of geographic information (Elwood, Goodchild & Sui 2012), inasmuch as most of these contents may embed a spatial reference, thanks to the availability of GPS and sensors in handheld devices and smartphones, as well as geo-browsers or location-based social networks, which are used for production.

This novel type of geographic information is commonly referred to as Volunteered Geographic Information, emphasizing the role of users which act as volunteer sensors to collect and contribute to this data (Goodchild 2007). However, the information produced and shared through social networks, namely Social Media Geographic Information (SMGI) (Campagna 2014), may be considered a deviation from the traditional VGI nature, since the collaborative collection and the diffusion of geographic information are not the main purposes of users (Stefanidis, Crooks & Radzikowski 2013). Despite an implicit nature of SMGI for the geographic dissemination, this kind of information, coupled with traditional VGI, proved to be useful in different application domains such as environmental monitoring, crisis management (Poser & Dransch 2010), as well as urban planning (Frias-Martinez et al. 2012). Indeed, VGI and SMGI may represent a valuable complement to traditional official information, supplying insights on users' perceptions and needs, opinions on places, as well as information about daily events, in (near) real-time, so potentially contributing to faster decisions.

In the urban and regional planning domain, VGI and SMGI may play an important role to support (1) analysis, (2) design and (3) decision-making,

fostering innovations in planning methodologies, inasmuch the majority of information required and used in practices is mainly spatial. As a matter of fact, this innovative wealth of GI may integrate the current availability of official digital information with pluralist knowledge from local communities that is usually neglected in practice, paving the way for innovative analytic scenarios.

Currently in Europe, a wealth of official digital geographic information was made available since 2007 by the implementation of the Directive 2007/02/CE (INSPIRE), which fostered developments in Spatial Data Infrastructures (SDI) among European member states. This process is favoring the access and the reuse of available official digital information, namely A-GI, produced by Public Authorities. This way, planners, analysts and professionals may access A-GI according to common technology, data formats, and policy standards. The integration of available official information with SMGI may further improve this potential, enriching the official datasets with information regarding not only geographic facts but also insights on people's perceptions and feelings both in space and time.

Nevertheless, the opportunities for the use of SMGI in spatial planning as affordable and potentially boundless sources of information have to deal with diverse challenges related to data management, data quality and data analysis. Indeed, the traditional spatial analysis methodologies and techniques may not be fully suitable to tackle the complexity of this information that exhibits Big Data nature for its modes of production and consumption (Caverlee 2010). Hence, in spite of several applications, related to different application domains and built upon the integration of SMGI and VGI in recent years, the lack of common methods to deal with these issues still requires the development of novel analytical frameworks in order to fully exploit the SMGI potential for analysis, design and decision making.

In the light of the above premises, this contribution presents a review of the authors' research results on the integration and use of A-GI and SMGI, aiming at developing valuable tools and methodologies for spatial planning. The remainder of the contribution is articulated as follows. The next section briefly discusses the distinctive features of SMGI, focusing on its main issues and opportunities for analysis. Section 3 introduces an original tool, developed by the authors and called Spatext, which enables the seamless collection, management and analysis of SMGI from multiple social media in a GIS environment. In section 4 a novel approach to SMGI analytics is proposed concerning a case study related to urban planning. Finally, section 5 draws conclusions and summarizes the discussion about the opportunities and the open issues of SMGI for urban and regional planning.

### **Issues and opportunities of SMGI**

The wealth of georeferenced user-generated contents regarding facts, opinions, and concerns of users, freely accessible through the internet by social media

Application Programming Interfaces (APIs), may affect current practices in urban and regional planning, giving opportunities for real-time monitoring of needs, thoughts and trends of local communities. However, the current public accessibility to SMGI is rather limited (Lazer et al. 2009), and common methods to manage, process and exploit these resources in practices still lack. The main hurdles limiting a wider use of SMGI may be found both in the shortage of user-friendly tools to collect and to manage huge data volumes and in the particular data structure of this information, which is burdensome to analyze by traditional methods. While the former challenge is starting to be addressed by new approaches typical of computational social science, an emerging field that aims to develop methodologies to address the complexity of Big Data (Lazer et al. *ibidem*), the latter challenge might require a tuning of analytical methodologies to deal with the several facets of SMGI.

First of all, although SMGI may be potentially available through the internet from any social media APIs, each platform features specific characteristics for contents production and sharing; hence SMGI from different social media may embed different attributes, causing difficulties for data integration and analysis. Moreover, SMGI is usually broadcasted through the internet by coupling alphanumeric data and multimedia clips, which complicate the analysis by means of traditional query languages. Secondly, SMGI, as user-generated contents with an associated geospatial component, combines the spatial and the temporal dimension of geographic information with a third dimension, namely the user itself, therefore extending the range of available analytical methods with further opportunities, such as users' behavioral analysis, users' interests investigation, land segmentation and potentially any analysis based on space, time and user (Campagna *ibidem*). These analytical methods may represent a novel way to investigate facets of the social and cultural habits of local communities, but their implementation may represent a challenge, which requires the integration of traditional spatial analysis methods with expertise and contributions from various disciplines such as social sciences, linguistic, psychology and computer science (Stefanidis, Crooks & Radzikowski *ibidem*).

The requirements for new analytical tools to deal with SMGI, and the opportunities resulting from the inherent nature of this information, guide the development of an original user-friendly tool, called Spatext, which eases the collection of information from multiple social networks and the integration of the data in a GIS environment for analysis.

### **Spatext: the SPAtial-TEmporal-teXTual Suite**

Spatext is an add-in for the commercial software ESRI ArcMap© implemented in Python 2.7. It enables the contextual social media data collection, management, geocoding, as well as the spatial, temporal and textual analysis of SMGI. This SMGI Analytics suite includes a number of tools, which can be

used mainly to (1) retrieve social media data from social media (including Twitter, YouTube, Wikimapia, Instagram, Instagram Places, Foursquare and Panoramio); (2) geocode or georeference data; and (3) carry out integrated spatial, temporal and textual analyses. In addition, the number of analytical methods available in the tool is steadily increasing in order to include several clustering algorithms to enable user profiling, user movement analysis, user behavioral analysis and land use detection, to name a few. Indeed, the collection, management and geocoding functionalities may turn any social media content into a workable SMGI dataset, which may then be directly integrated with other spatial data and analyzed in a GIS environment with off-the-shelf instruments.

Spatext takes advantage of the available social media APIs to perform queries directly from the GIS interface, enabling the collection of multimedia information regarding different topics, time periods and geographic areas. This way, the extension of traditional GIS tools with Spatext tools may ease the integration of SMGI with authoritative data, in order to support analysis, design and decision-making in urban and regional planning. The tools included in Spatext are developed in order to deal with the aforementioned issues regarding the access, management and analysis of 'big data' and can be categorized in three different classes according to the specific function: (1) data collection, (2) data management and (3) data analysis. The first class includes user-friendly tools that enable the harvesting of information from several social networks through spatial, temporal or textual queries. These tools can facilitate the direct access to social networks APIs avoiding programming efforts. The second class provides tools developed to ease the management, the integration and the successive analysis in GIS environment of SMGI extracted from different sources. These tools aim to limit the issues regarding the management and conversion of SMGI originated from different sources, which may present different data structures and information. Finally, the third class contains tools designed for analyzing the spatial, temporal and user dimensions of this information, as well as, for enabling the investigation of embedded textual contents. At the time being, the Spatext suite is not available for download due to minor technical revisions ongoing on APIs access. An overview of the Spatext functionalities is presented in Table 1, where the main tools are classified and briefly described according to the specific class, while the Spatext architecture is shown in Figure 1.

In the next section, functionalities of the Spatext tool for SMGI analytics are demonstrated through a case study related to urban planning in the municipality of Iglesias in Sardinia, Italy. The case study proposes the analysis of Instagram SMGI coupled with A-GI from Sardinian Spatial Data Infrastructures (i.e. Sardegna Geoportale <http://www.sardegnaeoportale.it>) to investigate the geography of the municipality and to debate the potential opportunities emerging from the integration of implicit experiential knowledge with official information for urban planning practices.

<b>SPATEXT TOOLS</b>				
<b>CLASS 1 - 'DATA COLLECTION'</b>		<b>query parameters</b>		
		<b>space</b>	<b>time</b>	<b>keyword</b>
Instagram extractor	Extracts Instagram SMGI to shapefile	✓	✓	✓
YouTube extractor	Extracts YouTube SMGI to shapefile	✓	✓	✓
Instagram Places extractor	Extracts Instagram Places SMGI to shapefile	✓		
Twitter extractor	Extracts Twitter SMGI to shapefile	✓	✓	✓
WikiMapia extractor	Extracts Wikimapia SMGI to shapefile	✓		✓
Foursquare extractor	Extracts Foursquare SMGI to shapefile	✓		
Panoramio extractor	Extracts Panoramio SMGI to shapefile	✓		
<b>CLASS 2 - 'DATA MANAGEMENT'</b>		<b>function activation</b>		
		<b>manual</b>	<b>automatic</b>	
Geocoding tools	Geocode places/addresses from attribute strings	✓	✓	
Conversion tools	Convert SMGI attributes to table	✓	✓	
Decomposition tools	Slice SMGI dataset in multiple datasets	✓	✓	
<b>CLASS 3 - 'DATA ANALYSIS'</b>				
Temporal Reference	Enriches SMGI dataset by adding information on temporal periods (season, month, day, hour) for further investigation			
Temporal Trend	Analyzes the temporal trend of SMGI contribution providing a graphic report			
Clustering functions (DBSCAN)	Detect clusters of high density by running the Density-Based Spatial Clustering of Applications with Noise algorithm on SMGI dataset			
Textual tag-cloud	Performs a simple tag-cloud analysis on textual contents of SMGI dataset to detect topic of interest			

**Table 1:** Main functionalities available in Spatext tool.

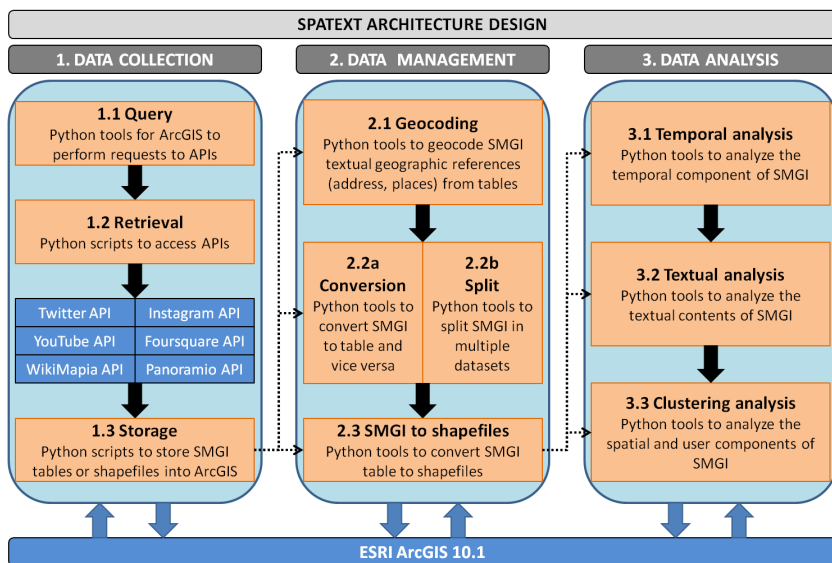


Figure 1: Spatext architecture design.

### Instagram SMGI analytics: an application in urban planning

In this section, an application of SMGI analytics is proposed through the analysis of Instagram contents in the urban environment of the Iglesias municipality in Sardinia, Italy. Nowadays, Instagram is one of the most popular online social networks worldwide, and it enables users to take, upload, edit and share photos with other members of the service through the platform itself, or other social media such as Facebook, Twitter, Tumblr, Foursquare and Flickr. Approximately 20 percent of the internet users aged 16 to 64 have an account on the service, and the trend is growing over last years. In addition, demographics of active Instagram users (GlobalWebIndex 2014) show a balanced percentage between male users (51%) and female users (49%), with a high percentage (41%) of users aged 16 to 24 that prevail over users aged 25 to 34 (35%), 35 to 44 (17%), 45 to 54 (6%) and 55 to 64 (2%). Statistics on the service stress also how a major part of active users (56%) appear to be into the middle quartile (33%) or top quartile (23%) of income. Among the features offered by Instagram, the geotag allows users to embed latitude and longitude of the place with the taken photos, therefore allowing to share the contents and the geographic reference through the internet according to own privacy settings. This capability plays a central role in considering Instagram contents as SMGI and permits the development of analysis to investigate spatial and temporal patterns within any geographic area where the service is available.

The case study concerning the Iglesias municipality (Italy) took advantage of the Instagram SMGI for a twofold purpose: (1) to explore the geography of the place through spatial and temporal patterns of the contributions, investigating

trends and areas of interest within the municipality, and (2) to identify and classify SMGI clusters, relying on the inherent spatial and temporal components, as well as by means of the integration with A-GI, in order to detect potential missing buildings in official datasets. The operational application of SMGI analytics on the case study of Iglesias municipality was carried out according to the following three main steps: (1) data collection, (2) analysis of spatial and temporal components, and (3) detection and classification of SMGI clusters, as explained in detail in the remainder of the contribution.

### *Data collection*

The data collection of SMGI from Instagram was conducted through the Spatext Instagram extractor tool by setting the spatial query parameter on the municipality of Iglesias and the temporal query parameter on a one year period (from 1 August 2013 to 1 August 2014). The extraction resulted in the collection of a one year sample of approximately 14,000 geotagged photos from 1,243 users for the study area. The tool automatically generated a point feature dataset, georeferencing each photo according to the geographic reference (latitude and longitude) embedded in the spatial metadata of the content, namely the geotag. Commonly, the geotag refers the GPS position of camera when the photo was taken; however, issues in connectivity may lead toward the lack of this information. In these cases, the Instagram service sets the geographic coordinates of the contents using the user's position during the upload. In addition to the geographic coordinates, the dataset includes several attributes, such as name of the place, if set by the user during upload, user name, user id, user picture URL, media URL, date of creation, number of comments, number of likes, tags and captions. These attributes are made available for any Instagram content if the user's privacy settings are public, offering opportunities for the development of several analysis in combination with other spatial data layers. Even though these pieces of information are publicly available, data were anonymized for privacy issues before any storage or processing for the study.

An exploratory analysis of the SMGI dataset showed a mean value of 11.22 photo/user, a modal value of 1.0 photo/user and a median value of 2.0 photo/user. Indeed, the 39.82% of users contributed with only 1 photo per year, the 32.74% contributed with 5 photo or more, while only the 4.34% of users posted 50 photos or more. Despite a different degree of participation by users, the dataset was investigated in order to identify potential commonalities among contributions in terms of areas of interest and urban dynamics.

### *Analysis of spatial and temporal components*

After the data collection, the spatial and temporal components of the SMGI dataset were investigated directly in GIS environment, in order to explore



potential patterns of interest in the area and local community dynamics. At this stage, the SMGI dataset was integrated with several official datasets from the regional spatial data infrastructure of Sardinia related to the Iglesias municipality such as settlements, roads network and buildings.

A simple investigation of the dataset spatial distribution showed a high concentration of the placemarks within the built environment, with approximately the 89% of the contents taken in residential or commercial and service areas. This value may depict the users' preference to employ the Instagram service in situations strictly related to their daily life within the city and might be considered a good starting point to investigate the dynamics in the municipality. The spatial distribution of the SMGI dataset is shown in Figure 2.

With the above considerations in mind, the temporal component of the SMGI dataset was investigated for different periods by searching potential peaks of interest, trends and dissimilarities in the use of Instagram by the users in Iglesias. The temporal analysis was performed investigating seasons, months, days of the week and hours of the day, disclosing interesting patterns. The results of temporal analysis showed how SMGI was increasingly produced and shared by users during the spring (30.9%) and summer (33.3%) in opposition to winter (19.1%) and autumn (16.7%); and this phenomenon was also evident in month distribution where July presented the highest percentage of produced contents (13%) and November the lowest one (5%).

The analysis of daily distribution provided more balanced results, with a slightly higher percentage of contents produced during weekends (Saturday and Sunday). Finally, the analysis of daily hours trend allowed identifying two main peaks of interest for both workdays (Monday to Friday) and weekends (Saturday to Sunday). The peaks were identified during the periods 14:00-15:00 and 21:00-22:00 for workdays, and the periods 14:00-15:00 and 20:00-21:00 for weekends, probably identifying meals or pause times. In contrast, the period 05:00-06:00 showed the lowest percentage of produced contents both for the workdays and the weekends. In spite of similar temporal peaks, the workdays and weekends trends exposed a few differences, which might be considered to be a descriptor of the typical cultural behaviors of inhabitants or a sort of cultural footprint of the place. This assumption may be corroborated by the results of a similar study conducted on Instagram datasets by Silva et al. (2013), which demonstrated how workdays and weekends temporal patterns were similar for cities of the same country, but showed major differences among cities in different countries. The results of temporal analysis for the different periods are provided in Figure 3.

### *Detection and classification of SMGI clusters*

The results of spatial and temporal investigations led towards the development of further analysis to investigate the geography and the urban dynamics of the municipality. Especially the major density of SMGI in the built environment

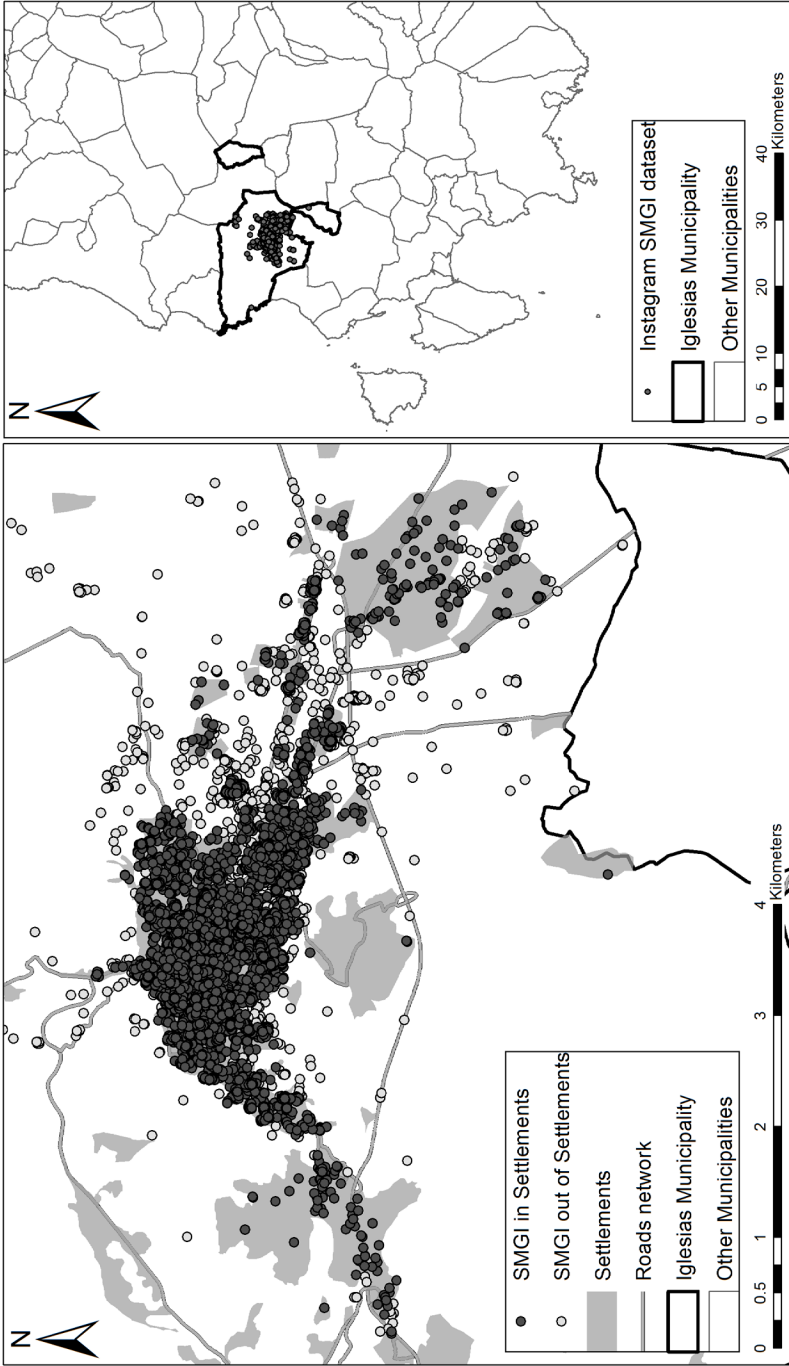


Figure 2: Spatial distribution of Instagram SMGI in Iglesias municipality.

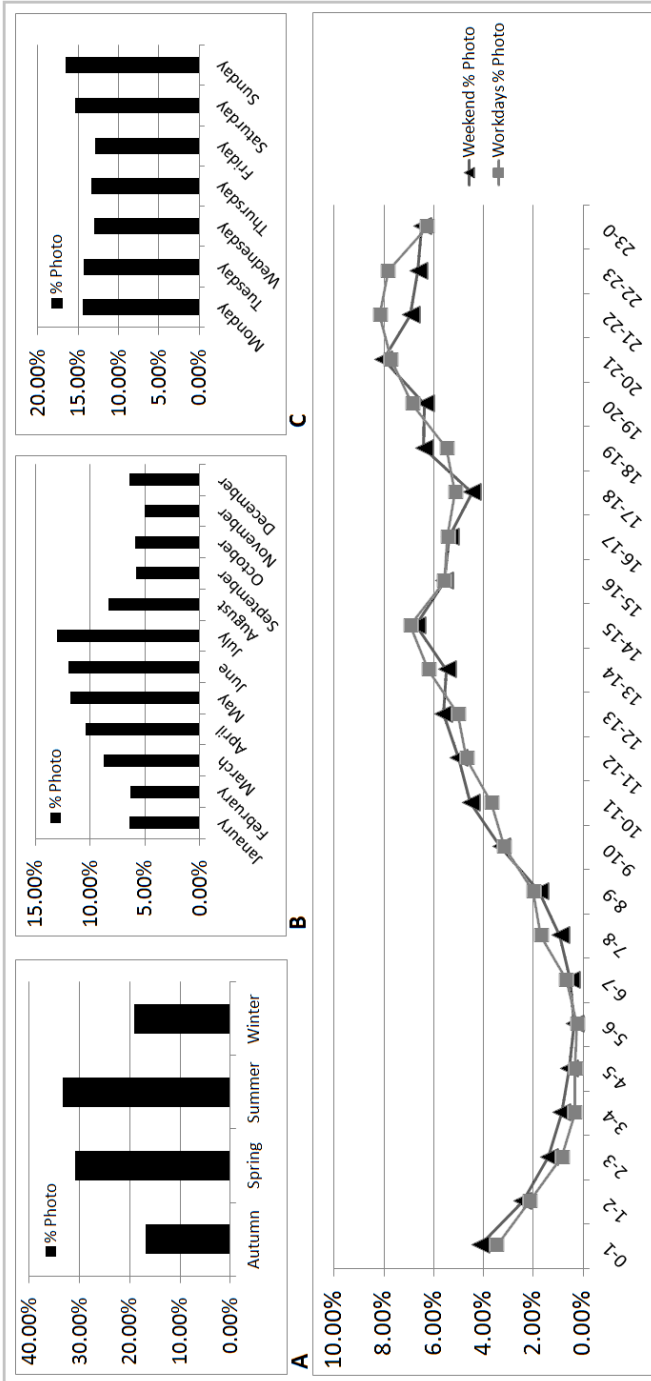


Figure 3: Temporal distribution of Instagram SMGI dataset: (A) season, (B) month, (C) day of the week, (D) hours trend during weekends and workdays.

fostered the development of analytical methods to identify, classify and interpret the users' interest toward certain specific spaces. For this purpose, the Density-Based Spatial Clustering of Applications with Noise algorithm or DBSCAN (Ester et al. 1996) and a slightly modified version called Feature-Based DBSCAN (FB-DBSCAN) were integrated in Spatext, and were used to compute clusters based on the spatial density of points.

The DBSCAN algorithm offers major advantages with respect to other clustering algorithms; firstly it is not necessary to know a priori the number of clusters, which also may differ in size and shape. Secondly, it works using two parameters exclusively: the epsilon (*eps*) that is the maximum threshold distance for including points in the same cluster, and the minimum number of points (*min\_pts*) that is required to define a cluster. In the study, the goal of the clustering analysis was the identification of the places that attracted the interest of the local community, which may be measured in terms of high density of contributions. Nevertheless, operatively there was no opportunity to establish the preferable value of *eps* and *min\_pts* before the computation, therefore the DBSCAN tool was applied iteratively on the SMGI dataset for different measures of the parameters in order to evaluate different results of the clustering. The assessment of clustering results led toward the identification of the following values, which proved to be the most suitable for the purpose of the study: *eps* = 20 meters and *min\_pts* = 5. Indeed, this *eps* value, or threshold distance, was able to cover the dimension of a medium-sized fabric, while the *min\_pts* value was set to 5 as a compromise value to avoid false positive in clusters detection and, at the same time, to prevent the dismissal of clusters with a modest participation of users. The results of clustering analysis with the above set of values enabled the identification of 290 clusters within the urban area of Iglesias, with a major concentration near the city center. In addition, two large clusters with an area greater than 50,000 square meters emerged from the analysis, identifying the areas attracting the highest interest by users within the urban context. These areas concerned both the historic centre of Iglesias and several service and public space areas. A closer look to the clusters showed that the top cluster contained the historic Cathedral of Santa Chiara, the main avenue for leisure and night life of the municipality, two of the main squares of Iglesias, as well as the train station area. At the same time, the bottom cluster contained several areas related to medical services, leisure, nightlife as well as the public park of the municipality.

Along the same vein, the FB-DBSCAN tool was used on the SMGI dataset in order to detect the places of major interest for each user. In fact, the FB-DBSCAN algorithm processes the dataset after performing a selection for attribute on the sample, in this case the users. This way, the algorithm computes clusters by processing only points related to a specific user for each iteration, offering opportunities to develop more specific analysis on the users' behavior. The analysis through FB-DBSCAN with the parameters *eps* = 20 meters and *min\_pts* = 5 identified 368 clusters concerning 266 users. In this case the

number of identified clusters was higher than the one in the previous analysis, but the clusters' sizes were notably smaller, identifying specific places or fabrics within the municipality. The results of the clustering analysis performed by DBSCAN and FB-DBSCAN are shown in Figure 4.

Each cluster identified through the FB-DBSCAN tool belonged to the contributions of a single user, and could be considered representative of a specific use regarding residence, work or leisure activities. The current use of a cluster may be discovered by analyzing several parameters related to spatial and temporal characteristics, as well as by integrating further spatial information. The aim of the study was the identification of not mapped buildings in the official information; therefore the latest official buildings dataset from the Regional SDI was integrated. This official dataset was selected in order to check the consistency of the clusters' location with the urban fabrics and to ease the identification of suitable parameters to detect residential clusters. As a matter of fact, the clusters related to a specific land use, in this case residential use, may expose similar patterns for certain characteristics such as number of intersections among clusters, temporal span among contributions, number of contributions and density of contributions, to name few, paving the way to the identification of common patterns for classification. In the study, six different parameters were selected with regards to the cluster itself and to the contributions, as described in Table 2.

The values of the six parameters were estimated for each cluster, and several combinations of the values were iteratively evaluated to identify exclusively the residential clusters. The following set of values resulted as the most suitable to classify a cluster as residential in the study area: *Cluster Centroid* and *Contributions Centroid* had to be 1 (yes), while *Number of Contributions* and *Time Span Among Photos* had to present the highest values among clusters of the same user, or the values had to be higher than 10 and 30, respectively. Finally, *Cluster Intersections* had to be equal or lower than 2, while *Cluster Density* had to be higher than 4. The above parameters allowed the identification of 47 residential clusters, which were confirmed by an overlay analysis with satellite imagery in GIS environment. Furthermore, the used parameters avoided potential biases caused by temporary phenomena such as massive tourists' presence or extremely popular events thanks to the threshold interval set for the parameter *Time Span Among Photos*, which considered only time periods equal or higher than 30 days to classify a cluster as residential. Afterwards, the same set of parameters was used to identify potential missing buildings in the official dataset by setting to 0 (no) the values of *Cluster Centroid* and *Contributions Centroid*, while leaving unchanged the other parameters values. Indeed, the values of *Number of Contributions*, *Cluster Intersections*, *Time Span Among Photos* and *Cluster Density* were considered as a sort of residential parcels footprint among clusters and were used for the investigation.

The analysis identified 40 clusters, which were then visually assessed through satellite imagery to confirm the presence of not mapped buildings in A-GI. The visual assessment allowed the detection of 9 not mapped buildings; at the same

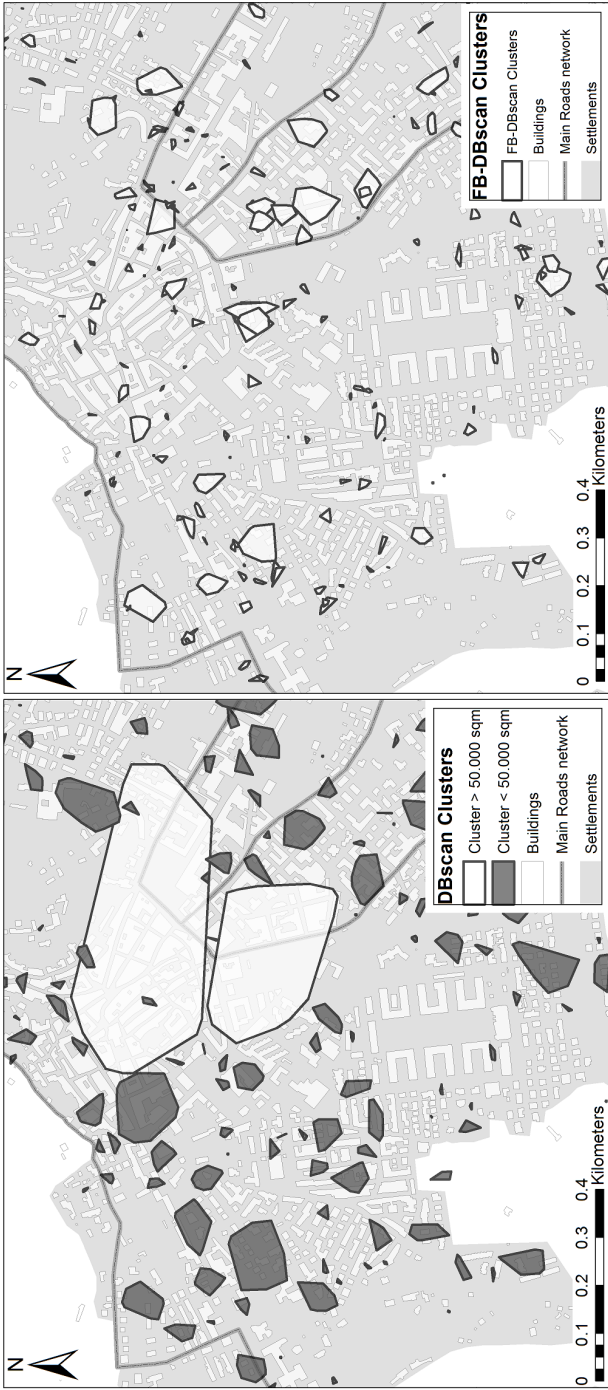


Figure 4: Clustering analysis of SMGI dataset: DBSCAN results (left), FB-DBSCAN results (right).

Parameters	Description	Units of measure
Cluster Centroid	The overlap of the cluster's centroid with an official building footprint is estimated	Boolean
Contributions Centroid	The overlap of the cluster's contributions centroid with an official building footprint is estimated	Boolean
Number of Contributions	The total number of contributions contained in the cluster is estimated	Number of contributions
Cluster Intersections	The total number of intersections between the cluster's shape with other clusters	Number of intersections
Cluster Density	The ratio between the cluster's area and the number of contained contributions	Square meters
Time Span Among Photos	The time passed between the first contribution and the last one in the cluster	Days

**Table 2:** Parameters used to identify residential clusters.

time the other 31 clusters were confirmed as residential areas, but the buildings were already mapped in A-GI. This issue can be explained by the lack of tolerance during the estimation of *Cluster Centroid* and *Contributions Centroid* values with the official buildings dataset. An example of the analysis results is provided in Figure 5, where six different clusters (i.e. A, B, C, D, E and F), their barycenter, the existing buildings footprints from the official dataset, the main roads network, and the Instagram SMGI dataset are shown.

In this example, the manual investigation through the Google Maps satellite image enabled the detection of two buildings which were not mapped in official dataset, namely cluster B and D. At the same time, the visual assessment confirmed the building presence in cluster A, C, E and F.

This example demonstrates the potentialities of Instagram SMGI to elicit information related to geography of places, and also shows how this information may be potentially used as a support for the update and the integration of official datasets.

## Conclusion

The results of the proposed study offer an overview of potential uses of SMGI for integrating and updating the available official information, as well as for obtaining information about the physical geography of places in the domain of spatial planning analysis. Currently, the wealth of information enclosed in SMGI may be used to investigate the concerns and the attentions of people toward places and also their behaviors and movements in space and time. These opportunities arise from the increasing availability of SMGI produced





**Figure 5:** Results of clusters investigation in Iglesias municipality.

through several social networks, which may be considered as affordable and potentially boundless sources of near real-time information about any topic. Hence, the collection of SMGI and the integration with official dataset may represent a valid support for analysis, design and decision-making, offering a pluralist perspective from local communities to enhance methodologies and practices in urban and regional planning.

Nevertheless, despite the several opportunities for analysis, it is important to be aware that the SMGI datasets should be not considered representative of the whole local community. The social network services are used differently by diverse segments of the population, that are the users of the service itself, and the preferences and cultural biases of these groups highly affect the phenomena under observation in SMGI, raising issues about the data representativeness. Furthermore, as for the subject of the proposed case study, namely the study of geography of a place by Instagram, the social platforms used to collect SMGI suffer of a different degree of penetration worldwide according to users' preference, limiting de facto the analysis opportunities only for areas where the services are available. In the future, a wider diffusion may occur to this respect as suggested by the current social network growth trends, but definitely for the time being both SMGI and A-GI show different diffusion rates in diverse regions and countries worldwide. Therefore, different analytical approaches



based on several platforms may be required in order to investigate the local contexts appropriately. Much more research is needed to assess the full potential of SMGI and several issues should be addressed regarding data quality and representativeness; however, current results disclose challenging research opportunities, which may lead to advances in spatial planning methodologies and practices, as well as in other domains.

### Acknowledgments

The work presented in this chapter was developed by the authors within the research project “Efficacia ed efficienza della governance paesaggistica e territoriale in Sardegna: il ruolo della VAS e delle IDT” [Efficacy and efficiency of landscape and environmental management in Sardinia: the role of SEA and of SDI] CUP: J81J11001420007 funded by the Autonomous Region of Sardinia under the Regional Law n° 7/2007 “Promozione della ricerca scientifica e dell’innovazione tecnologica in Sardegna”.

### References

- Campagna, M. 2014. The geographic turn in Social Media: opportunities for spatial planning and Geodesign. In: Murgante, B et al *Computational Science and Its Applications – ICCSA 2014*. Springer International Publishing. pp. 598–610. DOI: [http://dx.doi.org/10.1007/978-3-319-09129-7\\_43](http://dx.doi.org/10.1007/978-3-319-09129-7_43)
- Caverlee, J. 2010. A few thoughts on the computational perspective. In: *Presented during the Specialist Meeting on Spatio-Temporal Constraints on Social Networks*. Santa Barbara, CA, USA on December 2010.
- Elwood, S., Goodchild, M. F., & Sui, D. Z. 2012. Researching Volunteered Geographic Information: Spatial Data, Geographic Research, and New Social Practice. *Annals of the Association of American Geographers*, 102(3): 571–590. DOI: <http://dx.doi.org/10.1080/00045608.2011.595657>
- Ester, M., Kriegel, H. P., Sander, J., & Xu, X. 1996. A density-based algorithm for discovering clusters in large spatial databases with noise. In: Simoudis, E., et al (Eds.) *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining (KDD-96)*. AAAI Press, pp. 226–231. ISBN 1-57735-004-9.
- Frias-Martinez, V., Soto, V., Hohwald, H., & Frias-Martinez, E. 2012. Characterizing urban landscapes using geolocated Tweets. In: *Privacy, Security, Risk and Trust (PASSAT)*. 2012 International Conference on and 2012 International Confernece on Social Computing (SocialCom) IEEE, pp. 239–248.
- GlobalWebIndex. 2014. GWI Infographic: Instagram Users. Available at: <https://www.globalwebindex.net/blog/instagram-infographic> (Last accessed 08 April 2015).

- Goodchild, M. F. 2007 Citizens as Voluntary Sensors: Spatial Data Infrastructure in the World of Web 2.0. *International Journal of Spatial Data Infrastructures Research*, 2: 24–32.
- Lazer, D., Pentland, A., Adamic, L., Aral, S., Barabasi, A. L., Brewer, D., & Van Alstyne, M. 2009. Life in the network: the coming age of computational social science. *Science*, 323(5915): 721.
- Poser, K., & Dransch, D. 2010. Volunteered Geographic Information for disaster management with application to rapid flood damage estimation. *Geomatica*, 64(1): 89–98.
- Silva, T. H., de Melo, P. O. V., Almeida, J. M., Salles, J., & Loureiro, A. A. 2013. A picture of Instagram is worth more than a thousand words: Workload characterization and application. In: *International Conference on Distributed Computing in Sensor Systems (DCOSS) 2013*. IEEE, pp. 123–132.
- Stefanidis, A., Crooks, A., & Radzikowski, J. 2013. Harvesting ambient geospatial information from social media feeds. *GeoJournal*, 78(2): 319–338. DOI: <http://dx.doi.org/10.1007/s10708-011-9438-2>