UNIVERSITÀ DEGLI STUDI DI CAGLIARI

SCUOLA DI DOTTORATO DI INGEGNERIA CIVILE, AMBIENTALE ED ARCHITETTURA

DOTTORATO DI RICERCA IN INGEGNERIA DEL TERRITORIO

XXVIII CICLO

Social Media Geographic Information (SMGI): opportunities for spatial planning and governance

SETTORE SCIENTIFICO DISCIPLINARE DI AFFERENZA: ICAR/20

Doctoral Candidate: Pierangelo Massa

Doctoral Coordinator: Prof. Roberto Deidda

Supervisor: Prof. Michele Campagna

ESAME FINALE ANNO ACCADEMICO 2014 – 2015

ABSTRACT

The dissertation concerns the opportunities arising from the use of social media platforms as an information resource for supporting design, analysis and decision-making in spatial planning. The widespread diffusion of Web 2.0 technologies and tools such as geobrowsers, Application Programming Interfaces (API), GPS-enabled mobile devices, and recently Location-Based Social Networks are fostering the production, collection and sharing of georeferenced information by the Internet users, namely Volunteered Geographic Information (VGI) and Social Media Geographic Information (SMGI), which are not only related to measures of the geographical component, but also to user perceptions and opinions on places, localities and daily-routine events.

The wealth of VGI and SMGI freely available through the Internet may affect current practices in regional and urban planning, offering opportunities for real-time monitoring of needs, thoughts and trends of local communities. However, several hurdles related to data accessibility and management, as well as to knowledge extraction are limiting a wider use of SMGI in practice. In the light of the above premises, the research goal is to address the different aspects required for properly using VGI and SMGI within the urban and regional planning domains.

The methodological approach is developed following two main directions. First, the approach builds on the design and development of ad-hoc tools able to deal with the issues regarding the access, management and analysis of SMGI. Second, the dissertation formalizes a novel analytical framework, called SMGI Analytics, which enables the proficient use of this information in different planning scenarios. Several case studies are discussed in order to evaluate the value of both the developed tools and the proposed framework. Then, the SMGI Analytics framework is applied on a case study concerning the municipality of Cagliari in Sardinia (Italy) investigating and characterizing a specific public space.

Finally, the dissertation proposes a critical discussion about the developed tools and instruments' effectiveness for eliciting knowledge from SMGI. The discussion ends identifying the potentialities of obtained findings to address diverse questions related to spatial planning.

TABLE OF CONTENTS

CHAPTER 1

INTRODUCTION

1.1 Spatial planning in the digital age

Since the last decade, continuous advances in the Information and Communication Technologies (ICT), the Internet, and more recently, Web 2.0 technologies are starting to affect diverse domains of interest and are increasingly channeling digital Geographic Information (GI) into daily life of a growing number of users. This phenomenon represents a paradigmatic shift in how geographic information is produced and disseminated, as well as, in its contents and characteristics (Elwood et al., 2012), exploiting the new generation of digital GI and leading the renaissance of geographic information (Hudson-Smith and Crooks, 2008). In the spatial planning domain an unprecedented wealth of digital GI is made available to planners for supporting design, analysis and decision-making, disclosing great opportunities for innovations in methodologies. Major opportunities for innovation and development of methodologies emerge from the avalanche of "big" geographic information, which Web 2.0 technologies are making available to the wider public.

Spatial planning is a process mainly focused on the design and management of urban and natural environments, with the aim of increasing the well being of citizens (Frias-Martinez et al., 2012) and the correct functions of the territorial ecosystems. Among the different processes concerning urban planning, the characterization of urban land use and the identification of landmarks represent important processes, which require the provision of a large amount of accurate information on urban environment to planners, in order to develop sustainable decision-making processes and public policies. Commonly, this type of information is gathered by means of traditional methods of collection, such as direct observations or using residents/visitors' preference surveys (Jankowski et al., 2010). Nevertheless, these methods may present some limitations such as the survey running cost, the time consumption, as well as, the citizens' resilience to provide such information. Alternative methods based on the use of Geographic Information Systems (GIS) and remote sensing to analyze satellite imagery may provide useful information about land uses (Harris and Ventura, 1995), but these techniques may lose efficacy when dealing with requirements of updated or real-time information about urban environment.

Furthermore, novel theoretical and methodological approaches to spatial planning, such as Geodesign (Steinitz, 2012), and European policies on environmental assessment, namely Strategic Environmental Assessment (SEA) (Directive S.E.A., 2001), require a deep knowledge regarding both the geographic context and the cultural and social dynamics in order to develop processes in compliance to sustainable development principles. The Sustainable Development is depicted by the Rio Declaration (UNGA, 1992) as a complex concept, which includes several dimensions for achieving a democratic, environmental savvy, transparent and informed decision-making process. Similarly, the Agenda 21 program for Sustainable Development (UN, 1992) among its key concepts, suggests the development of participatory and social inclusive processes in order to design future development scenarios that are shared by the largest number of involved actors. Therefore, the implementation of transparent and participative bottom-up approaches, as well as, the integration of technical and official knowledge with public's experiential knowledge might inform the design of territorial development and eventually contribute to more sustainable development processes (Campagna, 2005). As a matter of fact, the concept of citizens' observatories for environmental protection is an issue also enclosed in the EU Framework Program for Research and Innovation Horizon 2020 (i.e. Call SC5-17-2015).

In addition, the proficient use of a broader and more pluralist knowledge about places may represent an opportunity of great potential to enrich 'smart city' strategies. Since last years, the concept of 'smart city' is playing an important role in the development of urban policies oriented toward sustainability and quality of life in cities by combining ICT solutions with investments in human and social capital (Toppeta, 2010; Caragliu et al., 2009). The elicitation of experiential knowledge and its integration with official information and sensor data infrastructures may foster the implementation of strategies actually informed by the local communities' requirements and opinions according to a bottom-up approach. This way, the advanced technologies might permit innovative forms of communication, governance and organization near real-time for the community engagement in assessing and solving urban key problems (Batty et al., 2012).

In the light of these considerations, exploiting the wealth of digital GI, collected into new digital formats and made available by the ICTs and Web 2.0 services, may represent an opportunity for integrating official and experiential knowledge in spatial planning processes to support design, spatial analysis and decision-making.

First of all, since the late 1990s, advances in Spatial Data Infrastructures (SDI) enabled the access to digital data, produced and maintained by public or private organizations for institutional or business purposes. Currently in Europe, the implementation of the Directive 2007/02/CE, establishing a shared INfrastructure for SPatial InfoRmation in Europe (INSPIRE), is leading to the development of SDIs in Member States and regions, granting the public access and reuse of available official information, or Authoritative Geographic Information (A-GI), according to common data, technology, and policy standards. SDIs impact may provide beneficial results for public administrations, developers and planning practitioners, and is slowly bringing innovation into the planning practices (Campagna and Craglia, 2012). For example, in many regions in Europe, the regional SDIs already represent the de-jure technical platform for the development of regional and local planning processes, by means of supplied data and services.

Secondly, developments in connectivity, geobrowsers and mobile technologies, enabled by Web 2.0, allow citizens acting as volunteer sensors to provide GI real-time in a bottom-up fashion. This information encloses both expert knowledge from professionals and experiential knowledge from local communities, producing an enormous opportunity for enhancing the available knowledge base in urban and regional planning. Currently, this wealth of digital information, commonly referred to as Volunteered Geographic Information (VGI), emphasizing the users' voluntary effort to collect and contribute to this geographic data (Goodchild, 2007), may be easily collected, analyzed, understood and used to support informed decision-making processes. These opportunities may enable a transactive approach (Friedman, 1973) in planning practices and lead to democracy and sustainability oriented approaches in plan-making, according to a communicative process (Innes, 1995). In several countries worldwide, the use of VGI is easing and fostering participatory processes in local planning for countries where authoritative data sources are absent at large scale (Rambaldi et al., 2006). In addition, VGI is becoming a main source of information in different application domains such as environmental monitoring, crisis management (Poser and Dransch, 2010), as well as, spatial planning.

More recently, the widespread diffusion of social media applications is encouraging the diffusion of geo-referenced multimedia (Sui and Goodchild, 2011), or Social Media Geographic Information (SMGI) (Campagna, 2014 A), over the global Internet. Users may easily access information and also be the producers and broadcasters of personal geo-referenced contents on location-based social networks. These capabilities have overtaken past barriers in data communication, and are disclosing innovative opportunities for disseminating and gathering geographic information among million of users, fostering the foreseen media convergence with GIS (Sui and Goodchild, ibidem). The social media contents, or SMGI, may be considered an innovative 'Big Data' source (Caverlee, 2010), and the traditional spatial analysis methodologies and techniques may not be fully suitable to tackle the complexity of this information and exploit the full knowledge potential (Campagna, ibidem A). However, the petabytes of freely and publicly available information offered through social networks might be detected and successively used to develop advanced analyses with the aim of eliciting knowledge for decision-making support (Zin et al., 2013).

The integration of VGI and SMGI with A-GI may spread novel analytical opportunities in spatial planning, with regards not only to quantitative measures of geography but also to qualitative information related to users perceptions and opinions on places, social and cultural behaviors, urban dynamics and daily-routine events in space and time (Campagna et al., 2013).

1.2 From social media to Big Data analytics

The widespread diffusion of Web 2.0 technology and tools such as geobrowsers, Application Programming Interfaces (API), GPS-enabled mobile devices, and recently location-based social networks are fostering the production, collection and sharing of georeferenced information by the Internet users, namely VGI and SMGI, which are not only related to measures of geographical component, but also to user perceptions and opinions on places, localities and daily-routine events (Campagna et al., 2013). An important measure of this avalanche of information can be easily appraised by available statistics on major social platforms (100 Social Networking Statistics & Facts, 2012). In addition to spatial data, VGI and SMGI offer important information about the human component behind the geographic phenomena (Coleman et al., 2009). Indeed, VGI and SMGI provide quantitative spatial data along with a set of qualitative multimedia attributes, which may enlarge the analytical dimensions to investigate. Petabytes of digital contents about any topic can be found through available internet services, fostering new opportunities for analysis and research.

Social media platforms can be considered as the natural evolution of microblogging systems, which offered opportunities for the management, the creation and the diffusion of information in a recursive cycle of production and consumption (Vieweg et al., 2010). The current capabilities of social networks lead to the inclusion of geographic information into users' generated contents, driving geography into their daily routines. As such, social networks could be considered as affordable and potentially boundless sources of information concerning daily life, specific events and also opinions, feelings and needs of users, related to geographical locations and facts. However, information provided by social platforms, should deal with several major issues, if intended to be used for meaningful analyses, such as: data reliability, data management and knowledge extraction. Social platforms offer different ways for management, sharing and extraction of contents, provoking a degree of uncertainty for the knowledge creation. Unlike traditional geographic information, SMGI either concerns dynamic processes on the Earth surface or users' related behaviors and perceptions of a specific time period and require advanced tools to support real-time monitoring, analysis and decision-making.

Moreover, several hurdles arise in finding suitable practices and procedures to manage the available avalanche of information. Advanced Big Data analysis could represent a suitable solution to extract and manage social media information. Indeed, a direct extraction of the content (what?) rather than the causality (why?) from data (Pohl and Pohl, 2013) could be performed, in order to avoid information volume issues and take advantage of current computing capabilities. In several domains, advanced Big Data analysis have been proposed and explored to manage the digital information wealth, supplying interesting results for several analysis purposes.

Similarly, issues related to knowledge extraction can be addressed by the application of Big Data analysis techniques, Social Computing analysis, and also by the integration of crowd-sourced with authoritative data. The current rise of social media and computational capabilities may allow the process of several multimedia contents (text, video, image, sound), disclosing innovative opportunities for the study of human beings and society (Manovich, 2011 B).

1.3 Research gaps

The wealth of georeferenced user-generated contents, namely VGI and SMGI, regarding facts, opinions, and concerns of users, freely accessible through the Internet by social media APIs, may affect current practices in regional and urban planning, giving opportunities for real-time monitoring of needs, thoughts and trends of local communities. Nonetheless, by now the accessibility to SMGI is rather restricted (Lazer et al. 2009), and a lack of common methods to manage, process and exploit these resources in practices is notable.

The main hurdles limiting a wider use of SMGI may be found both in the limited availability of user-friendly tools to collect and to manage huge data volumes and in the VGI and SMGI particular data structure, that is difficult to analyze by traditional spatial analysis methods. While the former challenge is starting to be addressed by new approaches, typical of computational social science, as an emerging field that aims to develop methodologies to deal with the complexity of big data (Lazer et al. ibidem), the latter challenge might require a tuning of analytical methodologies to deal with the many features 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 could embed different information as attributes, causing difficulties for integration and analysis. Moreover, SMGI is usually broadcasted through the Internet by coupling alphanumeric data and multimedia clips, making it cumbersome to analyze this information by means of traditional query languages only.

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, thus extending the range of available analytical methods with further opportunities, such as: user behavioral analysis, user interests investigation, land segmentation, and potentially any analysis based on space, time and user (Campagna, ibidem A). These analytical methods may represent an opportunity 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 et al., 2013).

1.4 Aims and research questions

In the light of the above premises, this thesis focuses on the opportunities that SMGI may disclose as a knowledge source in spatial planning, investigating different analytical approaches, which may be proficiently used to elicit knowledge from these sources. The research aims to investigate to what extent SMGI may disclose opportunities to integrate the availability of digital GI with the experiential knowledge, made available through social networks by users, in order to inform analysis, design and decision-making processes in spatial planning and governance.

As introduced earlier in this chapter, the integration of SMGI with A-GI may offer potentially boundless and affordable sources of information regarding not only geographic facts, but also local communities' perceptions, opinions and feelings in space and time. However, there is a lack of both user-friendly tools able to deal with the complex nature of SMGI and a shared analytical framework to collect, manage and process this information in practices.

Therefore, this research deals with several main questions in order to achieve the SMGI use in spatial planning:

1. What is the nature of SMGI?
2. What instruments are required to exploit SMGI in practices?
3. What analyses are required to take advantage of A-GI and SMGI?
4. What geographic scales are the most appropriate for using SMGI in practices?
5. How may the SMGI generated knowledge be used for supporting spatial planning?

With these considerations in mind, the thesis investigates the different aspects required for properly using VGI and SMGI within the urban and regional planning domains, with reference to selected approaches found in literature. Then, the thesis presents the development of an ad-hoc instrument, called SPATEXT (SPAtial-Temporal-tEXTual Suite), which enables the collection, the management, the integration with A-GI, and the contextual analysis of SMGI in a GIS environment. Several example case studies, related to spatial planning at different geographic scales, are conducted by means of SPATEXT in order to investigate the analytical opportunities disclosed by the different SMGI components (i.e. spatial, temporal, multimedia and user). These findings are then used to calibrate SPATEXT and to formalize a novel analytical framework, called SMGI Analytics, which may enable the proficient use of this information in different planning scenarios.

The methodological approach is implemented in order to achieve the following goals:

1) identify the particular components which define the SMGI nature;
2) develop user-friendly tools for easing the collection, the management, the integration and the analysis of SMGI and for testing their efficiency on example case studies;
3) formalize the SMGI Analytics framework based upon the integrated analysis of spatial, temporal, multimedia and user dimensions;
4) apply the SMGI Analytics on a complex case study strictly related to urban planning in Sardinia, by integrating SMGI from several social networks with A-GI in order to understand users' preferences, interests, opinions and requirements;

This way, the research results might allow the identification of potential strengths and weakness in urban environments relying on users' contributions, who freely shared information among different social media platforms. Moreover, the findings might be used to inform policies oriented to sustainable development, geodesign studies, and 'smart city' strategies. They could also be used for the design of specific interventions in the territory, oriented toward the local communities' quality of life and their real needs.

1.5 Research methodology

The methodological approach used for the research is developed following two main directions. First, the approach builds on the design and development of ad-hoc tools able to deal with the issues regarding the access, management and analysis of 'big data'. In this stage, a suite of tools, called SPATEXT, which eases the extraction, management and analysis of SMGI in GIS environment is developed. SPATEXT is a Python 2.7 add-in for the commercial software ESRI ArcMap© 10.1 and includes a set of tools, which can be used mainly to (1) retrieve social media data from social media, such as Twitter, YouTube, Wikimapia, Instagram, Foursquare and Panoramio; (2) geocode or georeference SMGI; and (3) carry out integrated spatial, temporal, and textual analyses.

Second, the approach takes advantage of the findings and assumptions provided by the study of Campagna (2014 A), where a framework for the analysis of SMGI in spatial planning is proposed, in order to assess the opportunities, arising from SMGI, for spatial planning and governance. Moreover, Campagna discusses the structure of the SMGI data model on the basis of a study carried on using a novel participative social media platform. This way, a set of potential analyses that an analyst could perform on this kind of data are found:

- Spatial analysis of user interests;
- Temporal analysis of user interests;
- Spatial Statistics of user preferences;

- Multimedia content analysis on texts, images, video, or audio;
- User behavioral analysis;
- Combination of two or more analyses (e.g. spatial-temporal analysis, temporal-behavior analysis).

The aim of the methodological approach, based on a set of consecutive steps, is twofold: on the one hand it is used in order to test the previous assumptions; on the other hand, it verifies the opportunities arising from each SMGI component and formalizes a SMGI Analytics framework able to support design, analysis and decision-making in spatial planning.

Operatively, the research methodology consists of the following steps:

1) Developing tools to extract, manage and analyze SMGI from different social networks;
2) Carrying on explorative analyses on example case studies to identify analytical opportunities;
3) Calibrating and developing new tools;
4) Formalizing SMGI Analytics;
5) Applying SMGI Analytics on a complex case study;
6) Analyzing findings and proposing solutions to inform decision processes in spatial planning.

A set of explorative analyses is conducted on selected SMGI datasets in order to evaluate spatial and temporal patterns of contributions, as well as to elicit useful information on users' perception and preferences from the wealth of unstructured textual contents embedded in this information. Moreover, a number of explorative analyses is carried on in order to evaluate the potentialities arising from the SMGI user dimension and to find suitable analytical methods for taking advantage of this dimension in practices. The underlying assumption is that different social networks may provide diverse types of information; therefore both the tools and the analytical framework should be calibrated in order to properly exploit SMGI for practices.

Several example case studies are developed in order to investigate new spatial planning opportunities, at the global, the regional and the local scale, arising from the spatial, temporal, multimedia and user dimensions of SMGI collected from three major social networks, namely Twitter, YouTube and Instagram Moreover, their findings are used to calibrate and develop further tools, as well as, to formalize the SMGI Analytics framework. The SMGI Analytics framework is then applied on a case study concerning a specific area in the municipality of Cagliari in Sardinia (Italy), which includes the main beach of the municipality, namely the Poetto, the regional natural park of Molentargius, as well as the neighborhood located between the beach and the park's salt marshes.

The research integrates both A-GI and SMGI from different social networks in order to study spatial and temporal patterns of contributions in that area, investigating how the location is used and perceived by local community and tourists. In order to investigate differences in appreciation by users, belonging to different categories, and to identify success points of interest in the studied area, a set of analyses are carried out, integrating spatial, temporal and user dimensions of SMGI with official information. The adopted methodology benefits from clustering methods and a geodemographic segmentation approach. Results provide useful hints, related to strengths and weaknesses of the area, in order to orient policy strategies and to integrate users' preferences and dynamics in decisions-making processes.

1.6 Outline of the dissertation

The thesis consists of 8 chapters.

Chapter 1 introduces the research context and focuses on the opportunities arising from 'Big Data' geographic information for spatial planning practices and methodologies. The emerging hurdles for a proper use of this information in practices guide the general aims of the research and the development of the methodological approach proposed in the study.

Chapter 2 is devoted to the review of the state of the art. The chapter analyzes the evolution of spatial planning investigating the role played by experiential knowledge in analysis and decision-making. Moreover, the chapter discusses the opportunities that digital GI, both authoritative and user-generated, may disclose for 'smart city' strategies and the novel approach of Geodesign.

Chapter 3 provides an overview of current digital GI sources for spatial planning, focusing on the increased advances fostered by ICTs and Web 2.0 technologies. The chapter concerns the evolution of both the Spatial Data Infrastructures (SDI), which enable the access and reuse of A-GI, and the novel sources of GI represented by the VGI and the SMGI, which are made freely available through the Internet.

Chapter 4 deals with the opportunities of VGI and SMGI for spatial planning, investigating the nature of this type of information and inquiring the potential hurdles related to trustworthiness, quality and reliability. Moreover, the chapter analyzes the use of SMGI in different application domains, such as disaster and emergency management, political science, social science and media studies, as well as, urban and regional planning. Finally, the chapter entails the emerging analytical opportunities for spatial planning methodologies.

Chapter 5 sets out the research objectives and describes the methodological approach used in the study. In order to address the issues limiting the use of SMGI in spatial planning practices, which concern both the lack of user-friendly tools and shared methodological approaches, the chapter presents an ad-hoc developed tool, called SPATEXT suite, able to ease the collection, management and analysis of this novel kind of information in GIS environment. Afterwards, the chapter proposes a novel analytical framework, called SMGI Analytics, which is structured in order to exploit the integration of A-GI and SMGI, easing the development of multi-dimensional analyses on the different SMGI dimensions. The SMGI Analytics aims to support analysis, design and decision-making in spatial planning processes, geodesign studies, as well as, to inform ‘smart city’ strategies with information about the users preferences and concerns.

Chapter 6 concerns the first SMGI Analytics applications in practices through several example case studies, conducted at different geographic scales using SMGI extracted from different social networks, namely Twitter, YouTube, Instagram and Foursquare. The chapter investigates both the SPATEXT suite capabilities for extracting, managing and analyzing SMGI, as well as, the opportunities arising from SMGI Analytics to gain useful insights about users preferences and behaviors. In addition, the obtained results are used to verify the methodological approach’s steps, which are then used for developing a more complex case study.

Chapter 7 discusses the SMGI Analytics framework application on an Instagram SMGI dataset in order to investigate and to characterize the public spaces of the Poetto and the Regional Park of Molentargius in the municipality of Cagliari, Sardinia (Italy). In this case, the SMGI user dimension plays a central role and the chapter introduces a geodemographic classification methodology for enabling the user profiling. The chapter builds upon all the introduced SMGI Analytics steps, which are operatively conducted until the development of several multi-dimensional analyses on the area. The final quantitative and qualitative results are able to provide insights on the concerns and the preferences expressed by the different identified population groups in the study area, disclosing opportunities to guide spatial governance and public policies. Actually, the proposed methodological approach may be applied in different geographic areas, in order to inform policies and practices, as well as, to foster the design of ‘smart city’ initiatives, easing the development of sustainable processes informed by the real requirements of local communities.

Finally, Chapter 8 summarizes the main findings of the thesis, identifying the answers to the research questions, and draws the conclusions to delineate potential future research streams concerning the use of the novel sources of digital GI in urban and regional planning.

CHAPTER 2

THE EVOLUTION OF SPATIAL PLANNING

2.1 Introduction

Nowadays, local community' involvement in decision making processes provides a novel kind of information, which may be integrated with the professional one, and successfully used in planning domain, contributing to take into account a citizens–oriented view on urban development issues (Davoudi, 2003). This chapter proposes an investigation of the planning theory's evolution in order to identify the role played by all actors in the planning arena, the way by which they participate and the modes by means of they affect the final decision, as well as, to inquiry the major or minor role played by participation and social inclusion into planning processes. The role of the technical rationality in spatial planning changes in respect to spatial and temporal contexts according to different political, social and cultural settings (Campagna, 2014 B), fostering changes in actors' roles both in participation (Arnstein, 1969) and in decision-making. Therefore, due to different temporal, social and political constraints, both the community and the stakeholders involved into planning processes may face hurdles in understanding the why and the how the final decisions are made. This phenomenon may represent a major issue when dealing with the sustainable development processes, inasmuch it encompasses several important dimensions such as responsibility, accountability, transparency, and eventually democracy of decision-making. Sustainable Development, as depicted by the Rio Declaration' principles (UNGA, 1992), may be considered a complex concept, which requires the inclusion of several dimensions in order to achieve democratic, environmentally savvy and based on informed decision-making development processes. Furthermore, in Agenda 21 (UN, 1992) two chapters concern the role played by the technology and scientific community in sustainability, and the ways by which information could be used for guiding decision-making processes.

The European policies on environmental impact assessment transpose into practices the requirements for democratic, informed and environmental savvy processes, as demonstrated by the Strategic Environmental Assessment (SEA) (Directive 2001/42/EC). SEA was introduced in 2001 with the aim of contributing to the integration of environmental considerations into plan-making processes in order to achieve high level of environmental protection according to sustainable development principles. Thus, the SEA concept entails a structured, rigorous, participative, open and transparent process, based upon environmental impact assessment, which is applied to plans and programs (Fisher, 2007). Nonetheless, many concerns regarding its real efficacy to actually inform decision-making in the regional and local land-use planning processes are raised (Sheate, 2004; Fisher, 2010), as well as many pitfalls are identified in its practical application (Parker, 2007). Consequently, research efforts in spatial planning are required in order to find methods to properly inform the design of territorial development by environmental considerations, and to increase the transparency and accountability of decision-making processes (Campagna, 2014 B).

In this respect, spatial planning theory may help to understand the role played by the community and the pluralist knowledge in different theoretical approaches of the discipline along its evolution to inform plan-making. In addition, raising the role of local community participation may be useful to foster the development of 'smart city' strategies informed by the local communities' needs and opinions in a bottom-up approach. As a matter of fact, the concept of smart city is strictly related to the concept of Sustainable Development and currently it plays a central role in the implementation of urban policies, as a way for combining innovative technological solutions to achieve sustainability and livability in cities (Toppeta, 2010). A smart city builds on investments in human and social capital, management of resources, transport infrastructures, and ICT to ensure sustainable development and quality of life (Caragliu et al., 2009). Although in the past ICT and Web 2.0 were considered important for smart initiatives (Wilson, 1997), and the Internet and broadband networks were seen as leading enablers of fundamental e-services (Kroes, 2010), the smart city strategies should be considered as the organic integration of different systems (Dirks and Keeling, 2009) or essential concurrent factors (Chourabi et al., 2012). For these reasons, local community opinions and experiential knowledge should be exploited to inform and to enrich the integrated system of a smart city.

Furthermore, local-community participation and experiential knowledge may represent important success factors for the development of collaborative decision-making processes, which are rooted in an innovative approach to planning and design, called Geodesign. Geodesign may be defined as an integrated process informed by environmental sustainability considerations, which involves a number of technical, political and social actors in a collaborative decision-making (Steinitz, 2012). This approach may constitute a promising solution to address current open issues in SEA, achieving informed design of territorial development, transparency and accountability in decision-making. The innovation of Geodesign springs from the awareness that ICT and Geographic Information Systems (GIS) may offer unprecedented power for an effective use of both scientific and societal knowledge in planning, design and spatial decision-making by taking advantage of an extensive use of digital spatial data, processing, and communication resources. Therefore, the current availability of innovative sources of digital geographic information, made available through the advances into the Internet and Web 2.0 technologies, could represent an important and innovative knowledge base to feed analysis, design and decision-making in spatial planning processes. Official sources of geographic knowledge, namely A-GI, and experiential sources of information provided by users and local communities, namely VGI and SMGI, might be integrated and used as knowledge basis to inform sustainable development processes in spatial planning and governance, offering a glimpse of opinions and needs of people actually affected by plans' decisions.

In the light of these considerations, this chapter briefly describes the evolution of the spatial planning theory, focusing on the role of participation to inform decision-making. Following, the main components related to smart city initiatives are described, with the aim of identifying the major elements that could be affected by the integration of experiential knowledge in analysis and decision-making. Finally, the Geodesign approach is described, focusing on the opportunities that new technologies and innovative sources of geographic information may disclose for spatial planning and governance.

2.2 The evolution of spatial planning theoretical positions

Since the industrial revolution, the evolution of spatial planning presented a sequence of different theoretical approaches, ranging from rational planning's "blueprint" models, wherein design, models and quantitative assessment fostered the development of built environment relying upon several keys spatial factors (Hall, 1988), until recent approaches belonging to communicative planning (Innes, 1995), wherein the communication and the actors' participation play a central role in the plan-making process. However, an unequivocal conception about the evolution of spatial planning theory does not exist (Khakee, 1998).

Several authors identify different theoretical approaches in their overview of spatial planning evolution. Healey et al. (1982) proposed seven theoretical positions starting from the rational-planning, which evolved in three positions called incrementalism, the advocacy planning and the implementation-oriented planning. On the other hand, other three positions were identified in the neo-marxist political economy approach, the new humanism and the pragmatism. The latter approaches were considered as opposite to rational-planning inasmuch these theoretical approaches discussed the rational theory. In an overview of the planning theory development, Friedmann (1987 B) identified four theoretical positions which were i) analysis, ii) social reform, iii) social mobilization and iv) social learning. The investigation of these positions allows the recognitions of similarities between policy analysis and social reform with the rational-planning and the incrementalism, as well as, the identification of commonalities between mobilization and social learning with the political economic approach and the new humanism. Friedmann himself may be considered the pioneer of the new humanism and as one of the authors who made important contribution to the social learning approach. In the social learning the planning process is depicted as a learning process, wherein the involved actors have to cooperate and to find common interests according to a transactive approach. The mutual learning is a fundamental step in this type of process and the knowledge required to guide plan' decisions relies on professional analyses and experience of local actors (Friedmann, 1973 A). Afterwards, Innes (1995) described the planning theory evolution between two specific paradigms: the rational-planning and the communicative planning. However, between these two paradigms, several theoretical approaches, concerning either the rational model or the communicative theory, exist.

Finally, Khakee (1998) took advantage of previous reviews and distinguished eight different theoretical positions, which are further investigated in order to study the role devoted to participation and experiential knowledge in the different planning processes. The eight identified theoretical positions were the following:

1) rational comprehensive planning;
2) incremental planning;
3) advocacy planning;
4) implementation-oriented planning;
5) strategic planning;
6) transactive planning;
7) negotiative planning
8) communicative planning.

The technical knowledge is integrated and enriched with the experiential knowledge when moving from the rational paradigm toward the communicative planning, wherein both the political stakeholders and the local communities play a major role in design and decision-making fostering mutual learning, democracy, accountability and transparency in planning processes (Friedmann, 1973; Innes, 1995; Khakee, ibidem). A brief description of each theoretical position is proposed in order to identify both the main involved actors and the role of participation and experiential knowledge in such planning positions.

The rational comprehensive approach assumes the planning as an ordered stepwise process that relies upon instrumental rationality. In this approach the decision makers define the goals while the planners should produce a set of alternatives able to fulfill in different ways most of the forecasted objectives. The main involved actors are politicians and professionals, while the knowledge base to guide the design is exclusively technical.

The incremental planning is mostly based on political factors, which determine the number of alternatives to be evaluated. In this case, the process is guided by political players that decide according to their own preferences, while the experiential knowledge of local community is usually neglected.

The advocacy planning springs from the assumption that many groups of actors may have different preferences and values. Therefore, notwithstanding several groups may influence the planning decisions process, many other groups require a professional voice to defend their interests (Davidoff, 1965). This approach recognizes the requirement of pluralism as an important factor in the society and starts to create a link between technical expertise and participatory democracy.

The implementation-oriented planning stresses the relationships existing between the making of goals and the policies with the implementation of the plan. This approach requires continuous formulations and revisions of the plan goals and policies in order to achieve the final result. For this reason, one fundamental step of the approach is the proper identification of involved stakeholders and the tailoring of plan objectives on their values in order to grant the implementation and the achievement of the forecasted results.

The strategic planning is mainly concerned on the uncertainty that may affect planning processes, fostering the development of modifiable alternative plans, which may be changed accordingly to the availability of new information. Indeed, especially social problems may provoke not only a quantitative uncertainty, but a qualitative one too (Dror, 1986), requiring adaptability and cyclic revisions in plan processes. In this approach, the politicians and the technicians represent the main actors, nonetheless the local community and the pluralism acquire a more important role to inform the planning process. The politicians foster the participation and choose the methods to integrate the knowledge base in a cyclic process of plan revision.

The transactive planning, or new humanism, takes advantage of the opportunities arising from the interactions among participants. The technical knowledge of professionals is integrated and enriched with the experiential knowledge of local community by involving in the discussion politicians, technicians and public in a mutual learning process. The transactive approach stresses the importance of participation, as well as the essential role of experiential knowledge, or local community values (Klosterman, 1983), to inform the planning knowledge base. As a matter of fact, planning is not exclusively a technocratic process (Thorgmorton, 1992) and a good planning process needs to engage local communities in its implementation (Zoppi, 2012; Wates, 2014) in order to deliver local benefits, as well as to be sustained over time (Leslie et al., 2007).

The negotiative planning is based on the collaboration between public authorities and market actors, in order to promote the development of a common planning project. In this approach, the public authorities foster the participation of business to achieve the best possible result; however, the public is usually excluded from negotiations in a such kind of process.

Finally, the communicative planning is an approach based upon the communication among involved actors, interlacing explanatory and normative aspects (Innes, 1995), in order to achieve inclusiveness, democracy, transparency and pluralism. This approach stresses the major importance of participation, pluralism and experiential knowledge for a profitable plan-making. In literature Healey (1993) distinguishes the ten main components that distinguish the communicative planning. First of all it is an (1) interactive approach, which relies on real world knowledge to establish actions and decisions. Then, it involves (2) discussion and dialogue among actors, eliciting the different interests and values and focusing (3) on the different emerged problems. In addition, (4) the planning process should rely not only upon policies, but also upon ‘arenas’ where the problems are debated and resolved. Therefore, (5) both the technical and the experiential knowledge are required for a (6) critical understanding of the problems. Such a result may be achieved exclusively through (7) the inclusiveness and the democratic pluralism fostered by (8) a mutual learning among actors. The interactions and the dialogues among actors should be based (9) on a clear communication, which (10) does not aspire to rational goals but instead tries to guide the process, which may be changed if necessary.

This brief review of theoretical positions in spatial planning focuses the different importance that experiential knowledge and participation may play in planning processes. Moving along the theoretical positions, from the rational comprehensive planning towards the communicative planning, the pluralism, the democracy, and the experiential knowledge gain major importance for proper decisions during planning processes. Nevertheless, the above review of planning approaches does not take into account considerations about the nature of information that is used to inform or enrich the knowledge base for planning. A reflection about the nature of information is fundamental when dealing with the recent approach of Geodesign, which identifies in digital geographic information the main component for the creation of a shared knowledge base among involved actors. Moreover, the Geodesign introduces the use of advanced methods applied through technological instruments in order to achieve democracy, transparency, accountability and participation along the plan process development. These considerations expand upon the opportunities that new sources of digital GI may disclose to inform geodesign approaches , as well as, 'smart city' initiatives, oriented towards the real requirements and values of local communities in the current digital age.

2.3 An overview on 'smart city' strategies

The real requirements, needs and values of local communities, as well as, the experiential knowledge that local actors could provide, may represent an important knowledge base for planning processes, affecting the current practices in design, analysis and decision-making, and informing 'smart city' strategies. Recently, the label 'smart city' emerged as a broad term to identify several strategies for solving problems generated by rapid urbanization and population growth in cities. A smart city strategy builds on the central role of the Internet and Web 2.0 to deal with several societal challenges, as well as urban welfare, societal participation, environmental sustainability and quality of life (Schaffers et al., 2011). In literature, several definitions of smart city can be found, concerning diverse elements that should be considered for the success of such strategies. According to Hall (2000), a smart city should monitor and enhance the condition of its infrastructures, plan activities and increase the offer of services to its citizens. ICTs and Web 2.0 should be considered fundamental elements to integrate, connect and make efficient the global system of infrastructures and services (Washburn et Sindhu, 2009), or to improve livability and sustainability in the urban systems (Toppeta, 2010). Moreover, the physical, IT, social, and business infrastructures are seen as fundamental components to foster the intelligence of the city (Harrison et al., 2010). Technology is fundamental to achieve 'smart cities' as sources of spatial enablement for citizens, in order to improve access, sharing and integration of spatial data with services (Roche et al., 2012). Despite the increasing and common use of term ‘smart’, the concept underlying these strategies was investigated and considered from different perspectives in literature, offering several clues and challenges for further investigation.

The technological component offered by ICTs and Web 2.0 should not be considered the only key to the success of such strategies, but rather successful results of smart initiatives may depend on the integration of technological components with managerial, political and contextual dimensions of the city (Nam et Pardo, 2011). ICT may be merged, integrated and used to coordinate traditional infrastructures and services, easing the comprehension and analysis of urban complexity. At the same time, technologies may allow innovative forms of communication, governance and organization for the community engagement in evaluating and solving urban key problems (Batty et al., 2012). These dimensions may be further delineated by a comprehensive set of factors as claimed by Chourabi et al. (2012). In the study eight factors were considered the most important for shaping smart city strategies, namely:

1) management;
2) technology;
3) governance;
4) policy;
5) community;
6) economy;
7) physic infrastructure;
8) natural environment.

This set of components may be referred to the requirements for collaboration, networking and coordinate interaction, as argued by Adam (1996) with regards to e-governance initiatives’ success factors. A similar classification of elements was provided also by a study conducted on a number of smart initiatives worldwide (Lindskog, 2004), wherein the synergy among components was considered the key factor for a successful achievement.

Among the several factors, governance, policy and community, enclosed in the political dimension, may play an important and central role for a ‘smart’ development. Several stakeholders are involved in the implementation of smart city strategies and a deep exchange of information and tight relationships are required to avoid the failure of projects (Scholl et al., 2009). Policies establish laws and regulations and may supply the enabling conditions for technology integration in urban development, accordingly (Dawes and Pardo, 2002). People and communities address a critical role in the development of smart cities, due to the fact that these strategies affect directly the quality of citizens’ life. Furthermore, citizens and local communities play a central role during the development process of smart cities, providing their needs and opinions by means of participatory initiatives, which in turn may guarantee transparency, democracy and pluralism.

In the light of the above considerations, notwithstanding the different smart city definitions found in literature, the participation and the experiential knowledge of local actors and people should play a central role for tailoring successful smart city initiatives. Therefore, the wealth of digital geographic information freely available through the Internet, supplying insights about opinions, needs and perceptions of local communities (near) real-time may disclose valuable opportunities for driving not only spatial planning and governance practices but also smart city strategies.

2.4 The Geodesign approach

The term Geodesign is recently emerged among spatial planning and GIS scholars identifying an approach to planning and design deeply rooted in geographic analysis and able to inform collaborative decision-making, or as claimed more broadly by Steinitz (2012), it is “changing geography by design”. Geodesign may be defined as an integrated process informed by environmental sustainability appraisal, which aims to solve complex problems related to territorial and environmental issues, as well as, to social and economic matters (Dangermond, 2010). Geodesign operates as a design and planning method which interlaces the development of alternative design proposals with a seamless impacts simulation informed by geographic contexts and digital technology (Flaxman, 2010). Despite the innovation in the term Geodesign (Artz, 2010), this approach may be dated back in time. As a matter of fact, many of the concepts entailed by Geodesign may be rooted in the landscape architecture traditions and environmental planning, as well as, they may be retraced to the concept of Sustainable Development. From an operative perspective, this approach is concerned with manipulating the set of processes and resulting forms, which operate on the Earth’s surface, in order to achieve specific objectives. (Goodchild, 2010).

As an integrated and multidisciplinary process, Geodesign includes project conceptualization, knowledge building, analysis, alternative design, impact simulation and assessment, decision-making, collaboration and participation, involving political and social actors and relying on scientific geographic knowledge support. The main innovation in Geodesign may be found in the extensive use of digital spatial data, processing, and communication resources, such as ICTs and GIS, which may enable a more effective use of scientific and societal knowledge in planning, design and decision-making (Ervin, 2011). Indeed, as claimed by several scholars and industry experts, the current technology may be considered mature enough to exploit the ICTs support in the planning practices, overcoming the barriers which until now have limited de facto the new technologies’ use (Göçmen and Ventura, 2010). As a matter of fact, since the last decade a growing wealth of both authoritative and user generated spatial data resources has started to being freely available, slowly shaping into reality the concept of Digital Earth (Craglia et al., 2012). Currently, two major categories of spatial data resources may be considered suitable for Geodesign approaches, namely A-GI from Spatial Data Infrastructures (Nebert, 2004) and spatial User Generated Contents (UGC), commonly referred to as VGI (Goodchild, 2007). These two types of spatial information are notably different in nature, but they might foster advances in planning and design practices exploiting informed decision-making and eventually contributing to more sustainable development processes (Campagna, 2005). Moreover, in urban and regional planning legislation, as well as in environmental assessment regulations, digital spatial data analyses and representations are starting to become mandatory for practices. For example in Italy, new regional spatial planning regulations (Paolillo, 2009; Campagna and Craglia, 2012) in Lombardy and Tuscany, as well as the Regional Landscape Plan in Sardinia, require the use of regional Spatial Data Infrastructures data and services to develop and represent local land use plans in digital format. As a consequence, the adoption of innovative spatial resources and tools may represent a way to comply with regulations and foster innovation in planning practices oriented toward the sustainable development.

In order to apply Geodesign in practices, Steinitz (2012) proposed a complete methodological framework to implement the approach in urban and regional planning and design. The Geodesign Framework relies on six models that are implemented iteratively in order to design future developments alternatives and to identify potential consequences of these scenarios by means of territorial context description, dynamics analysis and impacts evaluation. The six different models, defining the core of the Geodesign Framework, are representation, process, evaluation, change, impact and decision. The logical structure of the Steinitz's Geodesign Framework is shown in Figure 1, identifying which question each model aims to answer during its implementation.

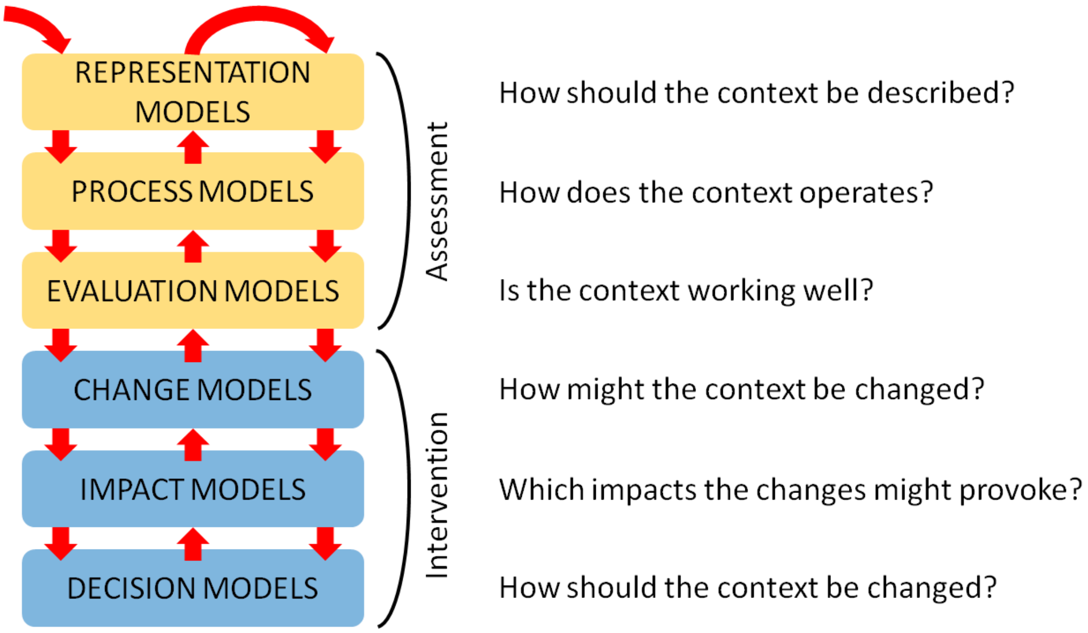


Figure 1. Geodesign Framework. Adapted by Steinitz (2012).

The first three models depict the current situation of the territorial context by describing (1) the environmental system and explaining (2) its evolution, mainly focusing on (3) opportunities and threats that may be devised. On the other hand, the last three models delineate potential alternative scenarios for (4) transformation, (5) assess potential beneficial or dangerous impacts on environmental and human systems, and eventually (6) support stakeholders during the decision-making process. The Geodesign Framework is shaped around three general questions, namely i) Why? ii) How? and iii) What, Where, When?, which affect the goal of each model in the framework, consequently.

A complete Geodesign study is implemented through three iterations along the six models. During the first iteration, the framework steps may be considered “data driven”, and the intent is to understand the geographic study area and the scope of the analysis, answering the Why? question. In the second iteration, the models are conceived in reverse order with the aim to specify how to carry out the study, and the process may be considered “decision driven”. Indeed, during this iteration, the geodesign team selects the methodologies and the tools which better fit the project purposes and answer the How? question. Finally, during the third iteration the designed methodology and models are carried out from the first to the last one, providing answers to the What, Where and When? questions according to a “data concerned” process.

The linearity along the iterations is not fundamental and a number of feedbacks or shortcuts might be required before the study completion. One of the main advantages for spatial planning, hailing from the Geodesign framework, may be found in the analysis capability to inform the design since the planning process’ early stage and to proceed enriching it during each loop until the final implementation. Therefore, Geodesign may constitute a promising approach to tackle current open issues in SEA, such as how to take advantage of societal knowledge and how to inform design alternatives, and eventually to foster innovation in urban and regional planning practices.

A Geodesign study pursues the participation and collaboration among involved stakeholders, technicians and public in order to deal with complex problem-solving and achieve sustainable plan solutions for the geographical context. In order to achieve these goals, one of the main innovations is the major role dedicated to digital spatial data and processing resources for fostering an efficient use of scientific and societal knowledge in planning processes. This latter type of information is usually unstructured and differs notably in nature from official information; however, it may be useful to offer knowledge related not only to geographic contexts, but also to users perceptions and opinions on places, localities and daily-routine events (Campagna et al., 2013). In the light of these considerations, Figure 2 shows the potential use of both official information and spatial UGC to inform models into a Geodesign study.

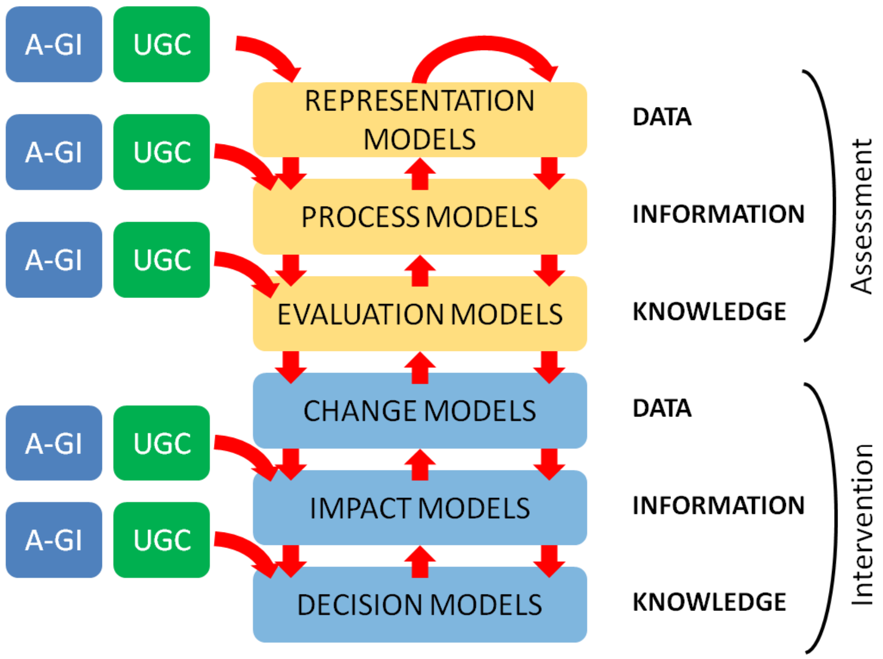


Figure 2. Integration of official and UGC into Geodesign framework. Adapted by Steinitz (2012).

The graphic provided in Figure 2 depicts how the use of official and experiential digital data might inform a geodesign study during the representation, process, evaluation, impact and decision models. As a matter of fact, the two different spatial data types may be used during the representation model in order to feed the proper abstraction and description of geographic context. The use of user generated contents related to a specific geographic area may enclose experiential information that is usually dismissed in official information, supporting a more pluralist vision of the geographic, social and cultural systems. On the one hand, the A-GI may offer official information about quantitative measurements, while on the other hand, spatial UGC, namely VGI and SMGI, may help in identifying particular social and cultural dynamics affecting the geographic context.

Following, A-GI and spatial UGC might be combined and processed during the process models in order to investigate how spatial phenomena evolve in time. As a matter of fact, VGI and SMGI provide updated and (near) real-time information, which may be used to feed predictive models and analyses aimed at identifying trends and phenomena affecting the area. Moreover, A-GI, VGI and SMGI might disclose notable opportunities to evaluate the current situation of the geographic context, providing further knowledge concerning the preferences and the social dynamics of users.

The integration of technical and experiential knowledge may represent a way to gain insights about social and cultural dynamics, which may help decision makers to promote a constructive dialogue about the future of places, proposing informed alternatives through the help of local community's experience (March, 1994). Commonly, the local knowledge of the residents is considered exclusively as an opinion in planning processes (Fischer, 2000; Rantanen and Kahila, 2009), but the technical knowledge, providing only a part of the required knowledge basis, may be not sufficient to properly guide the decision-making (Lindblom, 1990). Hence, the spatial UGC may be proficiently used to assess the alternatives impacts, supplying useful knowledge about the potentials and the risks of places (Rantanen and Kahila, ibidem). Finally, despite the experiential knowledge is difficult to articulate and convert into useful and explicit information (Nonaka and Takeuchi, 1995), it can be used through interaction among participants (Tsoukas, 2006). In this respect, planning should foster a communicative process, wherein the interlacing between expert and experiential knowledge is crucial (Khakee et al, 2000). The integration of A-GI, VGI and SMGI may support this process, providing knowledge about geographic and social context (Coburn, 2003), which may affect the decision-making processes. This way, both technical knowledge and experiential knowledge may be used in order to build a shared and sustainable development process for the territory among the different involved actors.

2.5 Discussion

This chapter discusses in the light of the evolution of spatial planning theory the roles that participation and pluralist knowledge play in planning processes related to different theoretical approaches. During the spatial planning evolution, several theoretical positions were found, ranging from the rational comprehensive planning, where the technical knowledge and rationality represent the unique required knowledge base, till the transactive approach (Friedmann, 1973) and communicative planning (Innes, 1995), wherein participation, communication and mutual learning among actors and experiential knowledge play a major role to guide the planning processes. The requirements for models evolution originate from the increasing issues emerging in the communication between professional or technical actors and the public (Khakee, 1998). Indeed, planners usually work with expert knowledge, while local communities are used to make value from their perceptions and personal knowledge. Important differences exist between these two types of knowledge, provoking misunderstandings and raising barriers in the communication among involved actors. The expert knowledge relies upon scientific methods and theoretical positions, whereas experiential knowledge is fed by daily life experience of citizens. Therefore, these two types of knowledge are difficult to integrate and the planning positions fostering communication, participation, democracy and pluralism might help in reducing obstacles (Khakee, 1998).

Moreover, the experiential knowledge may perform a central role in the development of smart city strategies oriented toward sustainable development. In fact, smart initiatives rely upon ICTs and Web 2.0 technologies to integrate, connect and make efficient the infrastructures and services systems (Washburn et Sindhu, 2009). Nonetheless, in order to improve liveability and sustainability in the urban systems, smart initiatives should depend upon the integration of technology with the political, social and cultural dimensions of the city (Nam et Pardo, 2011). Indeed, despite several stakeholders may be involved in the implementation of such strategies, the local communities address a central role, due to the direct impact of these strategies on their daily life. Thus, local communities' opinions, needs and perceptions may represent a valuable source of information to inform smart city initiatives with contextual experiential knowledge.

Finally, Geodesign is discussed as an innovative approach to spatial planning and design, that takes advantage of digital spatial data and geographic analysis to inform and develop collaborative decision-making processes (Steinitz, 2012). One of Geodesign notable innovation is the major role dedicated to digital spatial data, which may be used to inform the plan process with expert or experiential knowledge. Indeed, nowadays an unprecedented availability of digital geographic information is made available to planners to support design, analysis and decision-making. A geodesign study may take advantage of this information by means of advanced technological instruments, which should be able to collect, process and analyze this knowledge during the framework iterations, in order to achieve a more transparent, democratic and pluralistic planning process. This way, the study may be able to delineate alternative scenarios for transformation in the study area based upon technical, societal and cultural knowledge rooted in the geographic context, potentially satisfying the real requirements of involved people.

The digital spatial data made available for practices may be distinguished in two main categories: official information and spatial UGC, which present notably differences in production methods, analytical opportunities and sharing purposes. For this reason, there is an increased requirement of advanced instruments to collect, analyze and use proficiently the enclosed knowledge from these sources. In the following chapter a deeper investigation of the currently available digital spatial data sources is provided, focusing on the differences in nature and stressing the opportunities for spatial planning practices.

CHAPTER 3

SOURCES OF GEOGRAPHIC INFORMATION IN THE DIGITAL AGE

3.1 Introduction

In the last 20 years, continuous advances in the Information and Communication Technologies (ICT), the Internet, and more recently, Web 2.0, as well as the progress in mobile connectivity have fostered the extensive diffusion of new technologies into people's daily life of. Since January 2015, more than 3 billion people, almost 42% of the world's population, have access to the Internet through various devices, showing a growing trend of 21% since 2014. Moreover, about 3.6 billion people are mobile users, and currently the mobile's share of global web traffic reaches one-third of the whole net traffic (Kemp, 2015). At the same time, recent statistics show how the Internet users spend in average more than 4 hours per day using the net for different purposes, and 29% of population own an active social media account and use it more than 2 hours per day (GlobalWebIndex, 2014). In spite of a not evenly distributed access to the Internet and the new technologies worldwide, the growing trends of these numbers are representative of the current era of ICT and of the increased pervasiveness of technology in people's daily routine.

The widespread penetration worldwide of the Internet and Web 2.0 technologies are even strengthening the production, the sharing and the access of UGC (Krumm et al., 2008) among millions of users worldwide. Any multimedia content that is autonomously generated, shared and consumed by users through the Internet via diverse web platforms and social networks, can be considered an UGC. Currently, the wealth of user-generated contents, daily produced and accessed through the Internet, is transforming the Web in a potential and innovative source of digital geographic information (Elwood et al., 2012). Indeed, most of the UGC may embed a spatial reference, thanks to the global positioning system (GPS) and sensors availability in handheld devices, tablets, notebooks and smartphones, as well as, to the increasing number of geo-browsers, or location-based social platforms, which are used for production and dissemination. This phenomenon is fostering the convergence of social media and GIS (Sui and Goodchild, 2011), inasmuch GeoWeb platforms and location-based platforms are, increasingly utilized by users to provide geographic information and to facilitate interactions and constructive dialogues regarding places and social issues. In particular, the broadening of the GI collection, use and diffusion, from a small group of experts to potentially the whole community is triggering major changes to maps production and consumption (Engler et al. 2013), leading toward new scenarios of cartographic interactivity (Roth, 2013).

The diffusion of ICT and the increased availability of digital GI may foster noteworthy innovations in spatial planning methodologies and practices, enacting new ways of working, communicating and participating.

Since the 1960s, Geographic Information Systems started to be used into planning practices, enabling the use of digital information in replacement of traditional maps and analogue data. Nowadays, the consolidation of new digital data sources, produced by local communities, might improve the use of GIS into spatial government and disclose opportunities for a more collaborative, transparent and participatory decision-making in order to achieve sustainable development in planning practices at regional and local levels (Campagna, 2005). As a matter of fact, most of the required information to support analysis, design and decision-making processes is inherently spatial in nature, and these innovative GI may complement the current availability of official geographic information with the experiential knowledge of local communities.

3.2 Sources of Digital Geographic Information for spatial planning

In recent years, the advances in ICT, the Internet and Web 2.0 technologies are increasingly channeling digital GI into people's everyday life , easing the access to this information and disclosing innovative scenarios also for spatial planning. Indeed, the increased availability and accessibility of GI could foster notable innovations into urban and regional planning methodologies, allowing the integration of the official knowledge base with a broader and pluralistic vision of the territory and the urban systems, which is offered by the local communities contributions. Both expert knowledge from professionals and experiential knowledge from local communities may be easily collected, analyzed, and eventually used to support design and decision-making according to a transactive approach (Friedman, 1973), as well as, to facilitate the implementation of more sustainable and democratic communicative planning processes (Innes, 1995).

The major opportunities for spatial planning methodologies and practices hail from the unprecedented wealth of GI, currently available to the planners for supporting design, analysis and decision-making. First of all, since the late 1990s, the development in SDI enabled the access to digital GI produced and maintained by public or private institutions. More recently, the Directive 2007/02/CE (INSPIRE) is leading the development of SDI across European Member States and regions, easing the access and the reuse of A-GI to the wider public relying on common data, technology and standards.

Secondly, geo-browsers and location-based platforms are enabling the production, collection and diffusion of spatial UGC, where the community plays a more central role in production (Bruns , 2006). This innovative type of geographic information is commonly referred to as VGI, highlighting the role of users, which act as volunteer sensors to collect and to share this spatial data (Goodchild, 2007). The concept of VGI encompasses a wide range of activities and practices, which may provide pluralist sources of experiential knowledge from local communities and expert knowledge from professionals in a bottom-up approach for different application domains and planning processes.

Thirdly, the growing popularity of social media and Location-Based Social Networks (LBSN) is generating the diffusion of georeferenced multimedia (Sui et al., 2013), or SMGI (Campagna, 2014 A). SMGI may be easily accessed and shared by users, which seamlessly become even producers and consumers, or 'prosumers’ as claimed by Budhathoki et al. (2008), of geo-referenced contents. This type of information is a subset of VGI, but, since the collaborative and voluntary efforts for collecting and sharing GI are not the main purposes of users (Stefanidis et al., 2013), it may be considered a deviation from the traditional and common VGI vision,. Despite the implicit nature of SMGI with regard to geographic dissemination, this information, coupled with traditional VGI, may represent a valuable complement to traditional official information, offering insights on users’ perceptions and needs, opinions on places, as well as on daily-routine events, in (near) real-time, so contributing to faster decision-making responses and paving the way to innovative analytic scenarios in spatial planning. In the remainder of the chapter the three aforementioned digital GI sources are analyzed in more details.

3.3 Spatial Data Infrastructures and Authoritative Geographic Information (A-GI)

Since the 1960s, Geographic Information and spatial data started to play an important role in the geography domain (Foley, 2009) thanks to GIS developments and rapid advances in computational technologies. Indeed, GI and spatial data may be considered as fundamental in providing accurate information on cultural, social and natural resources, as well as, in describing patterns on the Earth surface (Maguire and Longley, 2005). Currently, GI is collected, processed and used for government, private sector and recreational uses, becoming a consumer good worldwide and representing a strategic resource, inasmuch the developed societies are becoming ever more spatially enabled (Williamson et al., 2004). In the 1980s, a number of national mapping and surveying agencies started to develop strategies and programs to facilitate the access to GI that were collected and used in practice. These strategies led toward the implementation of ad-hoc systems aimed to grant an easier access to databases of spatial data and geographic information, namely SDI.

The origin of the term SDI can be traced back to 1993, when the US National Research Council systems used the acronym to depict systems able to provide a standardized access to GI among other features (Mapping Science Committee, 1993). In 1994, the President Clinton’s executive order 12906 (Executive Order 12906, 1994) established the development of a national SDI (NSDI) in the United States of America, and quickly fostered the diffusion of SDIs initiatives worldwide (Budhathoki et al., 2008). Indeed, several countries started the development of SDIs in order to ease the access and sharing of spatial information between stakeholders involved in spatial governance and planning, granting the efficient use of geographic information to support decision-making processes within national boundaries (Crompvoets et al. 2004).

SDIs may be described as a set of policies, technologies and standards for easing the production, management and reuse of existent GI by spatial information users (Phillips et al., 1999), meanwhile reducing efforts and costs of dataset production through interoperability, sharing and easier access to geographic data (Davis Jr, 2009). This description is compliant with the one coined by the US Federal Geographic Data Committee (FGDC, 2004), which depicts SDIs as “the technology, policies, standards and human resources necessary to acquire, process, store, distribute and improve utilization of geospatial data”.

In spite of continuous advances in SDI development, since the mid 1990s, the implementation of these infrastructures experienced two different periods: first generation SDIs and second generation SDIs. The developments of first generation SDIs were mainly concerned on technical issues regarding the system architecture’s design and development in order to achieve specific purposes, such as: the geographic database completion and/or production, the economic development, the spatial government and the environmental sustainability (Masser, 1999). Each country involved in first generation SDIs initiatives promoted the SDI development based on national requirements and priorities, focusing more on the product, namely the spatial data, than on the real requirements of users and stakeholders. This product-centered vision was led by the national mapping agencies, namely the data producers, and limited de facto the private sector, civil society, and interested users involvement at large scale (Craglia and Annoni, 2007). The transition towards the second generation SDI started in 2000s when the leading countries changed their development strategies toward a more process-oriented or process-based approach (Rajabifard et al., 2003). Indeed, the second generation SDI initiatives are characterized by the central role of stakeholders within society, and the driving force for development shifts from data production to data use and users requirements. The distinguishing technological signal of the second generation SDIs is the web services introduction, which exploit a Service-Oriented Architecture (SOA) and act as fundamental components of the system in order to grant direct access to data and resources (Davis Jr, 2008), as well as, to satisfy the users requirements (Crompvoets et al., 2004) by means of added value services. Among these web services, the central role is played by the geoportal, which, as claimed by Tait (2005) represents “a web site considered to be an entry point to geographic contents on the web or, more simply, a web site where geographic content can be discovered”. Similarly, Maguire and Longley (2005) define a geoportal as “World Wide Web gateway that organizes content and services such as directories, search tools, community information, support resources, data and applications”. Therefore, a geoportal facilitates both the discovery of information through a repository of metadata, and the consequent access to these resources (Davis Jr, 2008) by making available to users a single entry point (Budhathoki et al., 2008). Important examples of geoportal may be found in the Geospatial One Stop (GOS) from the USA and in the EU-

Geoportal, which is a component of the INSPIRE project implemented by the INSPIRE Directive (Directive 2007/2/EC).

The Directive (2007/2/EC) establishes the legal framework for setting up and operating an INfrastructure for SPatial InfoRmation in the European Community (INSPIRE), which takes advantage of the SDI already existing in the European member states. The purpose of the INSPIRE Directive is to support the formulation, implementation, monitoring and evaluation of Community environmental policies (Craglia and Annoni, 2007). The Directive came in force in 2007, and since 2010 many European countries passed legislation introducing the INSPIRE requirements into national and regional laws, leading to the development of National and Regional SDIs initiatives according to common data, technology, and shared standards (Campagna and Craglia, 2012). The key elements of INSPIRE are i) 'metadata' to describe existing resources, easing the search and the access to information, ii) 'key spatial data themes and services' to provide a common framework for SDIs initiatives, granting interoperability and harmonization of resources, iii) 'network services and technologies' to enable the discovery, viewing, transforming and downloading functionalities, iv) 'agreements on sharing and access' to define common policies in member states, v) 'coordination and monitoring mechanism' to report on the Directive implementation and to evaluate its impacts and vi) 'process and procedures' to put in action the implementation process of the infrastructure.

The INSPIRE Directive builds upon a participative and collaborative process among official representatives from all Member States and public and private experts in environmental policy, with the aim of overcome five major barriers affecting the EU development (Craglia and Annoni, 2007) defined as i) the inconsistent data collection, ii) the inadequate documentation, iii) the incompatible datasets, iv) the incompatible geographic information initiatives and v) the data sharing barriers. The collaborative efforts, put in action by the involved stakeholders, pursue the definition of a common legislative framework to coordinate member states' SDIs initiatives according to a minimum set of common standards and processes. The final purposes leading the INSPIRE Infrastructure implementation (Craglia and Annoni, 2007) are the following:

1) the once collection and maintenance of spatial data across European countries;

2) the combination and reusability of spatial data from different European SDIs;

3) the sharing of spatial data between different levels of government;

4) the spatial data accessibility for extensive use into spatial and environmental governance;

5) the spatial data documentation, namely metadata catalogs, in order to expose data availability, fitness to purpose and conditions of use.

The INSPIRE Infrastructure is enabling the public access and reuse of available A-GI according to common data, technology, and policy standards with beneficial impacts for public administration, developers and planning practitioners (Campagna and Craglia, 2012). In addition, it is slowly bringing innovations into the planning practice, inasmuch in many regions in Europe, the regional SDIs represent currently the de-jure technical platform for the development of regional and local planning processes, by means of supplied data and services (Campagna and Craglia, ibidem). The term 'authoritative' refers to spatial data of SDIs that are produced by mapping agencies, experts, professionals and organizations for a specific mission or program and are compliant to well-defined institutional or legal frameworks (Ball, 2010; Goodchild and Glennon, 2010). The A-GI provision by trained experts grants the compliance with specific requirements and quality procedures, ensuring, at the same time, high degrees of accuracy and quality standards (Goodchild and Glennon, 2010; Elwood et al., 2012). Moreover, the official nature of SDIs spatial data is guaranteed by metadata, which describe several characteristics of this information such as: provenance, content, quality, accuracy, authorship, conditions of use, to name few (Nogueras-Iso et al., 2004).

In Sardinia, the Sardinian Regional SDI (SRSDI), namely the Regional Geographic Information System or "Sistema Informativo Territoriale Regionale – Infrastruttura di Dati Territoriali" (SITR-IDT) in italian, represents the technical platform to supply geographic data, services and technical resources to the public (Manigas et al., 2010). The SRSDI constitutes the regional geographic database wherein all Sardinian geospatial information are catalogued, maintained and made accessible according to the Legislative Decree 32/2010, which implements the INSPIRE Directive in Italy. Currently, the SRSDI performs a major role as official information system to support planning processes at the regional and the local level. In Chapter 6 and 7, spatial data from the SRSDI will be used and integrated with SMGI in several selected case studies in order to carry out analyses at the regional and the local scale in Sardinia.

3.4 Volunteered Geographic Information (VGI)

In recent years, continuous advances in Information and Communication Technologies (ICT), the Internet and new web services, loosely known as Web 2.0 technologies (O'Reilly, 2007), are strengthening the production, the sharing and the access of user-generated contents (Krumm, 2008) among millions of users worldwide. Most of these contents may embed a geospatial reference thanks to locationally-aware devices or location-based platforms used for production, leading toward the transformation of the Web in a potential innovative source of spatial data and information (Elwood et al., 2012).

As a matter of fact, advances in geospatial positioning, web mapping and collaboration technologies have surpassed the original vision of SDIs architects (Craglia et al., 2008), and the new technologies are notably altering the way by which citizens produce, access, use and share geographic information (Elwood, 2008 A).

Nowadays, a growing percentage of citizens may collect and share, actively or passively, geographic information concerning the environment and the urban contexts where they live, and also may enrich the spatial data with supplemental socio-cultural information and opinions through comments and tags (Castelein et al., 2010). This novel type of GI is commonly labeled as VGI, emphasizing the role of users which act as volunteer sensors to collect and contribute GI related to the geographic world (Goodchild, 2007; Elwood et al., 2012). The concept of VGI encompasses a wide range of activities and practices, which may provide pluralist sources of experiential knowledge from local communities and expert knowledge from professionals in a bottom-up approach for different application domains such as industry, government and academic research (Elwood, 2008 A). The VGI phenomenon is arising at fast rates worldwide facilitated by the technological advances and the paradigm of the social media movement (Elwood, 2010 B), as well as, by the aforementioned convergence of GIS and media (Sui and Goodchild, 2011). The ubiquitous diffusion of VGI on location-based platforms is notably augmenting and in certain cases is overcoming the information availability of several public and private sector web-based products (Haklay, 2010 A). In addition, VGI may foster the diffusion of participatory processes and may represent a useful source of information in different application domains to integrate or to stand in for authoritative data sources, when absent, at large scale. Indeed, many Internet users and citizens are granting the active and passive real-time collection of location-based information (Yin and Carswell, 2011), contributing to the construction of rich and increasingly complex spatial datasets to complement the availability of geographic information.

Despite its early stage, the theory and the practice of the VGI phenomenon may be considered parallel to other conceptual advances and research in the geographic science such as cybercartography and neogeography. The former can be defined as claimed by Taylor (2005) “the organization, presentation, analysis, and communication of spatially referenced information on a wide variety of topics interests and use to society in an interactive, dynamic, multimedia, multisensory, and multidisciplinary format”. Similarly, the latter, as defined by Turner (2006), indicates the “geographical techniques and tools used for personal activities or for utilization by a non-expert group of users; not formal or analytical”. Both the definitions emphasize the role of geospatial technologies and cartographic representations, which new web services may offer to citizens for production and dissemination of GI in a bottom-up approach. In contrast, the VGI definition concerns the data themselves, dismissing considerations on the technology, but in turn stressing the novel form of GI and the processes of voluntary production of information by users (Elwood, 2008 A).

The VGI phenomenon is generating unprecedented opportunities for enhancing democratic decision-making in spatial planning processes and is starting to be proficiently used in many applications domains. This information may be used to populate road atlases (e.g. Open Street Map - OSM), to support the reporting and relief coordination of natural disasters (Goodchild and Glennon, 2010), as well as, to integrate official information in the emergency management (Zook et al., 2010) and in the crisis management (Roche et al., 2013 B). VGI may also represent a valid support to respond and coordinate actions to deal with protests, riots and other political events (Liu and Palen, 2010), or to collect information related to environmental monitoring and spatial planning (Poser and Dransch, 2010), as well as to inform a knowledge basis for participatory processes in Citizen Science initiatives (Haklay, 2013 B) and in participatory planning (Knudsen and Kahila, 2012).

The spread of VGI initiatives worldwide is fostering a democratisation in geographic information production and consumption (Elwood, 2008 A), enabling the users' communities to share spatial data and to divide the production tasks collaboratively (Sunstein, 2009). In this case, the process of creation and consumption of GI is intertwined to the real needs of involved community rather than to the governmental priorities (Goodchild et al., 2007), allowing to highlight users' opinions, insights and interests. This democratic turn in spatial data production may spread opportunities to integrate, validate and potentially dispute the A-GI, providing evidences by means of experiential and local knowledge of individuals, which usually is neglected in official datasets (Engler et al., 2013).

Notwithstanding the common and shared vision of VGI phenomenon among academics, several contributions have been made in literature in order to offer a taxonomy of this information. As a matter of fact, most of the contents which are usually defined as VGI may exhibit different facets in terms of creation's methods and purposes. VGI might be distinguished between allocentric and egocentric information with regards to the production purpose (Elwood et al., 2012), emphasizing a community-oriented contribution or a self-interested one, respectively. Examples of allocentric VGI may include contributions to web-based map services such as Wikimapia and Open Street Map, to initiatives for coordinating disaster relief efforts, as well as, to citizen science initiatives. In contrast, egocentric VGI may provide information regarding the producer's position and movement, representing a means for personal social networking with others. In this case, examples may include georeferenced photographs shared through Flickr and Instagram, or the use of location-based social networks such as Facebook Places and Foursquare, to name a few. VGI taxonomy concerns also the nature of production, discerning between explicit and implicit contribution, as claimed by Craglia et al. (2012). Indeed, a VGI may be explicitly or implicitly created by a single user and this difference may be inferred from i) the format of the geographic content, ii) the production mode and iii) the will or desire of the producer for dissemination. An explicit VGI includes specific geographic coordinates, address or geographic reference to a building or a place, embedded in the contribution, while the production is the result of a voluntary effort of the user, who creates and shares the content with a specific purpose in mind for collaboration and sharing of GI. On the contrary, an implicit VGI provides geographical references expressed by means of natural language in tag and comments, or through metadata related to the location of production. In this case, the production effort accomplished by the user is not concerned to the dissemination of GI and it goes beyond the voluntary action intended for VGI (Coleman et al., 2009). Stefanidis et al. (2013) coined the term 'ambient geospatial information' to refer implicit VGI, meanwhile Jiang and McGill (2010) use the term 'opportunistic' sensing to indicate the harvesting of this type of information. However, the capability to collect implicit VGI without the awareness and permissions of producers may arise several risks for contributors such as identity theft, capture of tracking data and geo-surveillance (Scassa and Sattler, 2011; Thurm and Kane, 2010; Obermeyer, 2007, Sieber, 2007).

In the light of the above considerations, the voluntary effort of the users, underlying the production and the dissemination of geographical contents, may be considered one of the most important characteristics of VGI. Therefore, the motivations behind users' participation and collaboration represent a major topic of interest in the research on VGI phenomenon. Contributors may have altruistic purposes and participate in VGI initiatives for the creation of a common good by offering data and information to enhance the comprehension of particular regions and neighbourhoods (Coleman et al., 2009). At the same time, the voluntary motivations may lie in the opportunity to potentially affect decisions or to help local community, which may proficiently use this information (Gouveia and Fonseca, 2008). In addition, an egocentric behaviour may foster the production of voluntary contributions, which may be useful in turn for the personal reputation (Gouveia and Fonseca, 2008), the self-promotion (Goodchild, 2007), or for obtaining political outcomes and improving the social capital (Flanagin and Metzger, 2008; McDougall, 2009). Finally, besides the users motivation underlying the VGI contributions and albeit issues in data credibility, quality and resolution may persist, this information may offer important benefits in terms of support and information update thanks to the increasingly participation and collaboration of users with beneficial effects on data collection, logistics and social impacts (Gouveia and Fonseca, 2008).

3.5 Social Media Geographic Information (SMGI)

In recent years, Social Network Sites (SNS) and LBSN have gained a huge popularity over a short period of time among millions of users worldwide. Social networks may be considered as the natural evolution of microblogging systems (Vieweg et al., 2010) and may be defined as online communities of people who share common interests and activities (Miguéns et al., 2008) by means of specific web services, which offer several interaction' possibilities, ranging from a simple chat toward more complex forms of online communications, such as mobile connectivity, blogging, photo/video sharing and co-creation, modification and sharing of user-generated contents (Kietzman et al., 2011). More specifically, the SNS consist of web-based services, which allow users to create a public or semi-public profile within a bounded system, to articulate a list of other users or 'friends' with whom sharing a connection, and finally to view and to query their connections list and those made by others within the system (Boyd and Ellison, 2007). Most sites support the maintenance of already existent social networks in daily life, but others help strangers to connect according to shared interests, political views, or activities. Similarly, LBSNs allow users to communicate and to connect, but, thanks to the widespread adoption of location-based mobile devices and sensors, users may also share information about their location with their ‘friends’. The location, or namely the spatial dimension, is increasingly becoming a fundamental component of many online services, enabling people to share information about their geographic position and their movements with friends, meanwhile allowing companies to customize their services according to the users’ locations (Scellato et al., 2011). Thus, service providers and companies may access a valuable wealth of data on both the geographic location of users and the online connections among them, which disclose opportunities to investigate and exploit the spatial dimensions of the SNSs users, as well as to gain insights on real human socio-spatial behaviours (Scellato et al., 2011). The LBSNs appeal to a wide audience is driving geography into daily routines, fostering the convergence of GIS and social media as argued by Sui and Goodchild (2011) and enabling the sharing of knowledge not only about facts on the Earth surface but also about environmental, social and cultural phenomena. As such, LBSNs are transforming the Internet in an affordable and potentially boundless source of information about everyday life, events and also opinions (Gräbner et al., 2012), feelings and needs of users and local communities in space and time (Campagna, 2014 A).

The information contributed by the users to SNS and LBSN, namely SMGI as claimed by Campagna (2014 A), may be considered a special subset of Volunteered Geographic Information, inasmuch the voluntary production and sharing of GI is not the main purpose of the users. As a matter of fact, SMGI represents any multimedia content or information with explicit (i.e. coordinates) or implicit (i.e. place names or toponyms in natural language) geographic reference collected through SNS, LBSN or mobile applications. SMGI may include texts, images, videos, or audios in any combination, while the contents are aggregated in placemarks or georeferenced posts, commonly, featuring specific time reference and producer (Campagna, forthcoming). In literature, several terms are used to encompass this type of information such as egocentric VGI (Elwood et al., 2012), Ambient Geographic Information (AGI) (Stefanidis et al., 2013), implicit VGI (Craglia et al., 2012), involuntary Geographic Information (iVGI) (Fischer F., 2012) or User Generated Spatial Content (UGSC) (Antoniou, 2011).

In spite of a plethora of terms used to define the SMGI, the major interest raised by this information concerns the opportunity to study not only the geographic facts on the Earth surface but the people themselves, allowing investigations on movements, patterns and human behaviors in social and urban systems. Nevertheless, the information provided by SNSs and LBSNs should deal with several major issues for a proficient use regarding data reliability, data management and knowledge extraction. Indeed, social networks offer different ways for the contents’ management, sharing and extraction, provoking a degree of uncertainty for the knowledge processing. The reliability and quality offered by SMGI for research and practice are being discussed, and further efforts and investigations are required to define it (Jennex, 2010).

SMGI, in the same way as VGI, could be processed to elicit useful knowledge in relationship with specific degrees of uncertainty, in order to overtake credibility issues (Spinsanti and Osterman, 2013); however, several hurdles arise in finding suitable practices and procedures to manage the available wealth of information.

SMGI is usually accessible to the public through the Internet via API, giving opportunities for real-time monitoring of needs, thoughts and trends of local communities and consequently possibly affecting the way contemporary practices in urban and regional planning are developed. However, the public accessibility to SMGI is currently 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 and methods to collect and to manage huge data volumes and in the particular data structure of this information, which is difficult being analyzed by traditional methods. The former challenge is starting to be addressed by new approaches typical of the computational social science, an emerging field that aims to develop methodologies to deal with the complexity of Big Data (Lazer et al., 2009). On the one hand, advanced Big Data analysis may represent a suitable solution to extract and manage SMGI by a direct content extraction (what?) rather than the causality (why?) from data (Pohl et Pohl, 2013) in order to avoid information volume issues and take advantage of current computing capabilities. On the other hand, the challenge regarding SMGI data structure might require a tuning of analytical methodologies to deal with the several facets of this information.

Firstly, albeit SMGI may be potentially available through the Internet from any social media APIs, each platform presents specific features regarding contents production and sharing; hence SMGI from different social media could embed different information as attributes, causing difficulties for integration and analysis. Furthermore, SMGI is usually broadcasted through the Internet by coupling alphanumeric data and multimedia clips, making it burdensome to analyze this information by means of traditional query languages only.

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, thus extending the range of available analytical methods with further opportunities, such as user behavioral analysis, user interests investigation, land segmentation, and potentially any analysis based on space, time and user (Campagna, 2014 A). In order to better explain the particular characteristics of SMGI, Figure 3 shows a graphical representation of a SMGI placemark, stressing the available analytical dimensions and the potential embedded multimedia contents.

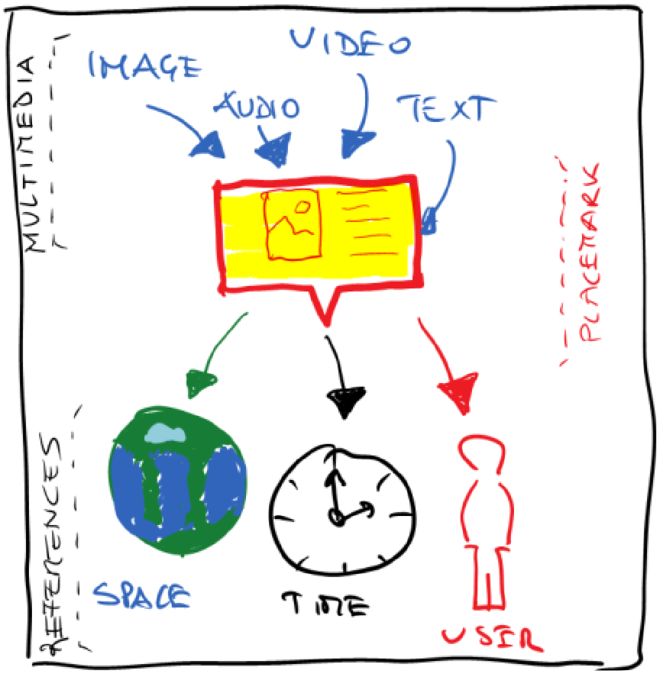


Figure 3. SMGI data model. Campagna (2014 A).

The range of different analytical methods to deal with SMGI may represent an opportunity 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 et al., 2013). Actually, issues related to the knowledge extraction have started to be addressed by the application of Data Mining techniques and Social Computing analysis, as well as by the integration of SMGI with authoritative data. Moreover, Spatial Data Mining and Geographic Knowledge Discovery are emerging research fields, which focus on the development of theory and methodologies to extract useful information and knowledge from complex spatial databases (Andrienko and Andrienko, 1999;. Miller and Han, 2009). The methods provided by these research fields are exploratory in nature and more inductive than traditional ones, including clustering, classification and association rules mining techniques, as well as, visual analytics (Miller and Han, 2009) in order to integrate and elicit useful knowledge from complex spatial dataset. In literature, several studies conducted through the SMGI analysis were found, embracing several fields of interest, such as: disaster events response, political events, media events, social studies, urban planning and tourism planning. Innovative instruments and analysis may be applied to detect events and information related to disasters by means of Twitter contents (Li et al., 2012 A) or Twitter and YouTube contents (Zin et al., 2013). Flickr tags may be used in semantic analysis for the development of social studies (Rattenbury et al., 2007), or for investigating people movement and landmarks preferences in urban environment (Jankowski et al., 2010). Along the same vein, the temporal component of Twitter SMGI may be used to investigate and validate urban land uses according to the real human dynamics (Frias-Martinez et al., 2012), while Booking.com and TripAdvisor SMGI may be used to investigate tourists' preferences patterns on destinations (Floris and Campagna, 2014).

In spite of difficulties in data access, management and analysis, the means of information available from SNSs and LBSNs about facts, opinions and feelings of users is starting to be proficiently used in several domains and it may represent a valid support for design, analysis and decision-making in spatial planning practices with a real-time monitoring of needs and requirements of local communities.

3.6 Discussion

This chapter discusses the current wealth of digital geographic information which is available to planners and practitioners thanks to advances in Information and Communication Technology, the Internet and the widespread diffusion of Web 2.0 technologies among million of users worldwide. In the last decade much progress has been made in technology to support creation and diffusion of digital geographic information, which represents the knowledge basis for several spatial planning practices. An unprecedented wealth of digital geographic information is available to support analysis, design and decision-making in spatial planning domain. Nowadays, planners may rely both on authorative and innovative contents and sources to search and collect suitable information for analysis, as such as: A-GI from SDI, VGI and SMGI.

Firstly, the implementation of the INSPIRE Directive (2007/02/CE) in Europe, has enabled developments in SDI, which are allowing the access and the reuse of A-GI according to common data, technology, and policy standards, with benefic impacts for public administration, planners, developers and consultants in planning (Campagna and Craglia, 2012). Secondly, developments in geobrowsers and mobile technologies have overtaken limits in acquisition and communication, enabling citizens to act as volunteer sensors (Goodchild, 2007) to crowdsource and broadcast GI in real-time with an innovative bottom-up fashion. Finally, the widespread diffusion of social media is fostering the diffusion of geo-referenced multimedia (Sui and Goodchild, 2011), or SMGI over the global Internet (Campagna, 2014 A). Users can easily access information and also be the producers and broadcasters of personal geo-referenced contents on location-based social networks.

Despite the potential of SMGI may be considered still limited for the public both in terms of accessibility and availability of analytical instruments (Lazer et al., 2009), the integration of A-GI, VGI and SMGI may ease a further step towards the next generation of geospatial intelligence and, may disclose innovative scenarios in spatial planning and smart city design and governance. Indeed, the combination of available A-GI with VGI and SMGI may enhance the potentialities of these sources for analysis and decision-making, complementing the official datasets by coupling authoritative information of public authorities with experiential knowledge of local communities. Actually, the traditional spatial analysis methodologies and techniques might be not adequate to manage and take advantage of their knowledge potential, therefore new methods and instruments for collecting information, managing huge data volumes, as well as eliciting reliable knowledge should be developed.

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 service users itself, and the preferences and cultural biases of these groups highly affect the phenomena under observation in SMGI, raising issues about the data representativeness. In addition, the SNSs and LBSNs used to contribute and to collect SMGI suffer of a different degree of penetration worldwide according to users' preference, limiting de facto the analytical opportunities to geographic areas where the services are available. In respect to this, a wider diffusion may occur as suggested by the current social network growth trends, but since the time being, different analytical approaches based on the use of several platforms may be required in order to investigate the local contexts properly.

CHAPTER 4

SOCIAL MEDIA AND VOLUNTEERED GI FOR SPATIAL PLANNING

4.1 Introduction

The current widespread diffusion of social media applications is producing an unprecedented availability of information regarding facts, events, users' opinions and feelings, over the global Internet. Examples of this information deluge may be easily depicted by the available statistics on different social platforms. Facebook social network exhibits over 1 billion of registered users, of which 552 million are daily active and spend over 6 hours monthly on the social platform. Similarly, Twitter, Google+, Linkedin, Instagram, Pinterest, Flickr and YouTube, to name few, exhibit considerable values in terms of registered users and contents production (100 Social Networking Statistics & Facts, 2012). Every day more than 500 million tweets are sent (Twitter Usage, 2015), over 3.5 billion queries per day are submitted on Google (Internet Live Stats, 2015), 70 million pictures per day are uploaded via Instagram with over 300 million active users monthly (Instagram Press, 2015), and 300 hours of video are uploaded on YouTube every minute (YouTube Press, 2015). Potentially, UGC (Bruns, 2006) about any topic could be readily found through the available Internet services, disclosing innovative opportunities for analysis and research. Furthermore, particular interest is raised by social networks contents that carry a geographic reference, commonly referred as VGI (Goodchild, 2007; Sui and Goodchild, 2011, albeit this type of information, for the particular modes and intents of production and sharing purposes, may be considered a particular subset of VGI, and more specific terms are gaining popularity such as iVGI (Fischer, 2012), AGI (Stefanidis et al., 2013), SMGI (Campagna, 2014 A).

The increased availability of SMGI over the global Internet is made available by the growing adoption of electronic devices equipped with GPS receivers (e.g. smartphones and tablets) and sensors in the past few years, which transformed most of the social network application into LBSN (Roick and Heuser, 2013). Nowadays, SMGI may play a major role in different application domains, since they may be proficiently used for investigating movement, patterns, events and users opinions in (near) real-time, so depicting a picture of what is happening in any potential place (de Albuquerque et al., 2015). Nevertheless, major issues arise into finding ways for managing the data volume and for eliciting useful knowledge from these sources. Indeed, due to the high velocity, sheer volume and particular structure of social media contents, significant challenges arise in how to deal with this 'Big Data' in order to extract exclusively pieces of relevant information and elicit knowledge. Unlike traditional data, SMGI refers to dynamic processes and requires new kind of tools to treat monitoring and decision-making in real-time concerning information that is continually changing, as well as, to link this crowd-sourced data with official information. In many situations, data obtained from SMGI should be integrated with authoritative geographic information in order to leverage the full potential from these sources. In addition, the quality and the reliability of SMGI for research and practices is actually being discussed and further investigations are required to establish the credibility's extent of this information.

4.2 SMGI reliability, quality and credibility

In recent years, the VGI phenomenon offered an alternative mechanism for the production, acquisition and compilation of geographic information regarding measurable elements on the Earth surface and environmental and socio-cultural phenomena happening on it. As such VGI offers notable advantages in terms of affordability and updating, but suffers from a general lack of quality assurance as highlighted by several authors in literature. Both VGI and its SMGI subset present an intrinsic heterogeneity in data (Girres and Touya, 2010; Haklay et al., 2010; van Exel et al., 2010), due to different participants' knowledge basis, to collaborative production processes and to specific systems that may be used for acquisition and production by the different participants. Moreover, unlike the traditional authoritative geographic information that these sources potentially complement, VGI and SMGI carry no assurance of quality, raising questions about the reliability, accuracy and credibility for practices. Nonetheless, the technological interoperability offered by SDIs might be proficiently used to integrate these sources with the official information (Craglia, 2007) and, at the same time, several authors are proposing different methodologies to assess of VGI and SMGI quality, fostering the integration with authoritative information.

The label 'authoritative' refers to geographic Information and spatial data produced by experts, professionals and mapping agencies, which operate under institutional or legal frameworks with a specific aim (Goodchild and Glennon, 2010; Ball, 2010). Therefore, the production of this type of information by highly trained experts complies with specific requirements in terms of quality and assurance procedures, guarantying the accuracy and the quality standards of the final product (Goodchild and Glennon, 2010; Elwood et al. 2012). Moreover, the authoritativeness of A-GI is assured by metadata that describe contents, quality, authorship, accuracy and use of this information conditions (Nogueras-Iso et al., 2004). On the contrary, VGI and SMGI are acquired, produced and shared by individuals' groups, which voluntary operate for a specific purpose or interest but without common practices in terms of quality and assurance procedures or in absence of adequate training.

In order to deal with the issues arising from the uncertainty underlying VGI and SMGI, several studies used different elements of spatial data quality such as: positional, thematic and temporal accuracies, logical consistency, and completeness (ISO/TC 211, 2002). The first developed studies concerned the investigation and comparison of OSM data, one of the most popular VGI project (Haklay and Weber, 2008), with A-GI from different countries worldwide. A number of studies assessed the positional accuracy and completeness of OSM' roads network (Haklay, 2010; Zielstra and Zipf, 2010), natural features (Mooney et al., 2010), as well as both the features (Girres and Touya, 2010). How the accuracy of the VGI position may be very high for manmade features rather than for natural features have been demonstrated by these studies (Bègin et al., 2013). Furthermore, the accuracy and the quality are proved to be high, even though the existence of semantic differences among features of voluntary and official datasets (Al-Bakri and Fairbairn, 2012). Nonetheless, a quality issue emerged in the datasets completeness, which concern places with low density population and consequent scarce number of contributions. Due to the particular nature of VGI and SMGI, new assessment methods should be used to integrate the traditional approaches and to take into account the heterogeneity nature and the collaborative production processes underlying this information (Van Exel et al., ibidem; Sieber, 2007). In this respect, Maue (2007) proposes the inclusion of information regarding authorship, reputation and knowledge basis of the author into metadata in order to validate the geographic information. Flanagin and Metzeger (2008) suggest the implementation of a validation system directly in the production interface, meanwhile Bishr and Mantelas (2008) proposed the implementation of automatic filters for assessing quality and reliability of information. More recent approaches suggested the integration and comparison of VGI and SMGI with official information from SDIs to automatically validate or reject the contributed information (Spinsanti and Osterman, 2013).

The aforementioned approaches for assessing VGI and SMGI, in spite of difference in methods, may be classified according to three main categories proposed by Goodchild and Li (2012):

1) the crowd-sourcing approach,

2) the social approach,

3) the geographic approach.

First of all, the crowd-sourcing approach benefits from the crowd, namely the people, in order to validate and correct the errors that an individual might make, thanks to the ability of a group. This approach relies on the Linus's Law (Raymond, 1999), formerly the capacity of the crowd to converge on the truth, which is used in the context of software engineering to identify and correct software' bugs. This principle might be proficiently used also to validate and assess VGI and SMGI, and several studies verified this principle (Latonero and Shklovski, 2010) focusing on the relationships between contributors' density and data quality (Napolitano and Mooney, 2012; Neis et al., 2011) or between the number of edits and features quality (Keßler et al., 2011; Mashhadi et al., 2012). However, when VGI and SMGI refer to a barely populated location or to an area affected by lack of interest, this approach may require a change of technologies.

Secondly, the social approach is based on the acknowledged reputation of the authors to assess the quality of the data they produce, leading toward the creation of a hierarchy of trusted individuals, who may act as moderators (Goodchild and Li, 2012). As a matter of fact, the tracking of individuals' contributions may allow the calculation of reliability metrics, providing trustworthy basis about the geographic information. Several studies rely on this approach using a data-centric evaluation based on the editing history of VGI contributions, exclusively (Keßler et al., 2011; Mooney and Corcoran, 2012), while other studies assess the contributions according to specific data and users characteristics (van Exel and Dias, 2011).

Finally, the geographic approach relies on a comparison of VGI or SMGI with the geographic knowledge, benefits from specific rules, which govern what may occur or not at certain locations (Goodchild and Li, 2012). As an example, the First Law of Geography by Tobler (1970) that says "all things are related, but nearby things are more related than distant things" is an efficient rule to assess the reliability of voluntary information. Indeed, a fact about a location should be consistent with both what is already known about that area and other similar acknowledged facts regarding the same location. However, the rules governing the geographic domain range from the very simple to the highly abstract, causing major challenges for the development of an adequate system for VGI and SMGI assessment.

At the current stage, VGI and SMGI may present notable advantages in terms of affordability and timely data, as well as in the capability to provide information that has never figured in mapping practices before. On the other hand, this type of information suffers from heterogeneity and incomplete coverage, failing to be considered as a proper and valid alternative source to A-GI, though it may play a major role in exploratory analysis and in integrating official datasets.

In literature several applications were found for different application domains (i.e. disaster events, political events, media events, social studies and urban planning), promoting methodologies and innovative tools to provide solutions for dealing with aforementioned issues.

4.3 The use of SMGI in application domains

The current wealth of affordable and timely information over the Internet regarding facts, events, opinions and feelings is disclosing enormous opportunities for the study of human beings and society (Manovich, 2011). The potential use of SMGI in different domains of interest is gaining popularity among scholars, professionals and researchers, which, in recent years, investigated, analyzed and verified the fit to purpose of this information for practices. Several studies deal with the use of in different fields, such as disaster and emergency management, analysis of political events, investigation of social dynamics, as well as, urban and regional planning. In addition, innovative methods and tools, which address the issues emerging from the inherent heterogeneity and 'big data' nature of SMGI. Although each study may provide differences in methods and tools, according both to different design purposes and specific domain of interest, several analogies may be found in analysis and technology. In the light of these premises, the next sections present a brief summary of several studies wherein SMGI plays a central role for analysis and decision-making.

4.3.1 Disaster and emergency management

A growing research body is concerned with the analysis of the potential use of social media contents, namely VGI and SMGI, for disaster and crisis management. Different methodologies and procedures dealing with SMGI for extracting useful knowledge to support analysis and decision-making are found in literature.

First of all, an application to retrieve disasters' eyewitness photos from Flickr is exposed. The application proposes a methodology based upon a qualitative study to investigate if and how users' activity on Flickr may evolve in case of notable disasters (Liu et al., 2008). By means of a qualitative study, the collected images regarding the hazard, the post-impact and the online convergence might be significant to disaster response efforts. In the study the photos of latter kind, especially, combining data from different sources, are found useful to create new overviews and overlapping maps, supplying spatial information on the event location and the potential relief resources. The study proposes an approach to capture SMGI, stressing the potential of this information for disaster response and recovery issues; however the analytical methodology relies upon a manual intervention for the identification and the extraction of useful data in order to implement a qualitative analysis.

Secondly, an application to enhance the situational awareness (SA), through analysis of Twitter posts during a disaster is reviewed (Vieweg et al., 2010). The aim of the study is the harvesting of real-time contents during a crisis event according to lifecycle production and consumption of information in microblogging. SMGI concerning two disaster events occurred in the 2009 in USA, relying on the Twitter API to perform the collection of data has been analyzed. The information extraction from tweets is based on well-defined terms related to each event, taking advantage of an initial investigation of the Twitter public stream. Afterwards, the geographic references of each tweet are obtained through a manual analysis of users locations in order to achieve a manageable dataset. The proposed analyses expose differences in behaviors of people between the warning phase, namely anticipatory awareness of a disaster, and the impact phase, or real-time awareness of the event. Moreover, the results exhibit a growing percentage of georeferenced tweets during the impact phase, suggesting the intent of users to supply more useful information to the online community. The application proposes several procedures and analysis for managing SMGI of Twitter, as well as, the use of advanced instruments and technology. Nevertheless, a manual intervention is even required for data collection and analysis.

Thirdly, another study presents an ad-hoc application to detect events related to disasters by Twitter contents. The Twitter Event Detection Analysis System (TEDAS) aims to detect and rank new events according to their importance, generating spatial and temporal patterns of the extracted data (Li et Al, 2012). The proposed application relies on Java, PHP and Application Programming Interfaces of Twitter and Google Maps to collect tweets and the related users' location, according to well-defined terms for research queries. The obtained results offer an overview of spatial and temporal patterns of the detected events in real-time directly. In this case the study takes advantage of more technologically advanced tools, exclusively, which limited manual harvests.

Similarly, another study presents an approach to categorize tweets by analyzing the Twitter broadcasted contents during the Black Saturday brush fires of 2009 in Australia (Sinnappan et al., 2010). Here the authors classified the tweets using a categorization scheme designed for disaster events, specifically, in order to evaluate the percentage of tweets providing potential useful information. The results, based on a sample of 1684 tweets, demonstrate how only 5% of SMGI contained proficient data regarding the disaster with directly actionable information, while the information remained may be classified as noise or not directly actionable.

Along the same vein, a study conducted by Starbird et al. (2010) analyzes data collected from a particular subset of users commenting on the Red River flood of 2009 in the United States and Canada, looking at the particular features of their discussion. The aim is the investigation of the Twitter usage patterns during the disasters, relying on the qualitative analysis of tweets, which contained several disaster-related terms, collected during the event period. The results conduct to the identification of two overlapping categories of useful tweets, namely generative and synthetic. The former category includes tweets that provide new information through the description of lived experiences and facts, while the latter contains tweets that merge a variety of information from external sources and spread this new information. In addition, the results demonstrate how major hurdles arise in finding original tweets, for representing less than 10% of the analyzed sample, as well as, tweets from users directly afflicted by the event that represent less than 2% of the sample.

Zin et al. (2013) introduce an application developed in order to extract visual and textual data from YouTube and Twitter respectively, for describing the situation awareness related to disasters). The study proposes and approach composed by several steps to analyze SMGI, focusing on location, network, contents and time. Statistical operations are carried out on collected data to rank the detected events according to the relative importance; however, the application requires an empirical procedure to a proper data management, leading toward differences in results if the events detection relies on textual SMGI or visual SMGI.

In their study, Spinsanti and Ostermann (2013) present the Geographic CONtext Analysis of Volunteered Geographic Information (GeoCONAVI), a prototype system which aims to retrieve, process, analyze and finally evaluate social media related forest fires contents. The goal of the approach is to evaluate the opportunities that VGI and SMGI may offer as trustworthy and actionable information during disaster events. In this approach the authors integrate the SMGI dataset with further official available information concerning the geographic context of the events, so enabling an immediate assessment of quality and reliability. Moreover, the approach builds on a set of spatial-temporal clusters techniques in order to support the scoring and the final validation of retrieved information.

In addition several studies are concerned on the capability to detect such kind of phenomena from the use of social networks in the domain of disaster and emergency management are carried out.. For example Sakaki et al. (2010) use the microblogging platform Twitter for event detection, which benefit from real-time production and consumption of information. A few semantic analyses are applied to Twitter SMGI in order to classify them into positive or negative classes via a support vector machine (SVM) approach. The aim of the study is to develop an earthquake reporting system, able to detect an event based on sensory observations, namely the tweets. In order to identify the earthquake events and establish the correct location, the reporting system relies on probabilistic models and location estimation methods.

Similarly to the previous study, Crooks et al. (2013) proposes an approach to analyze the spatial and temporal dimensions of Twitter contents related to a 5.8 magnitude earthquake, which occurred on August 2011 in the East Coast of the United States. The authors demonstrate how Twitter feeds may represent a hybrid form of a sensor system allowing the identification and localization of the event epicenter. Despite the limited quantity of extractable Twitter SMGI, which is only a 1% sample of the real volume, the study demonstrates how it is possible to detect the event location within 5% of time that is usually required by currently available systems. In addition, interesting spatial-temporal patterns, starting from the impact area firstly and then spreading over time across other locations, were identified from tweets origins. Results demonstrate that an early collection of Twitter SMGI may be used to provide a rapid approximation of the earthquake impact area.

Finally, a study proposed by de Albuquerque et al. (2015) recommends an approach for integrating social media contents with official information in order to deal with disaster events. The suggested approach aims to enhance the identification of relevant messages from Twitter relying upon the intertwined relations between SMGI and the contextual geographic features of the event, which may be derived from authoritative data, such as sensor data, hydrological data and digital elevation models. The study is carried out analyzing Twitter SMGI produced in June 2013 during the River Elbe Flood in Germany and applying statistical analysis for detecting general spatial patterns in the occurrence of event-related SMGI, associated with both proximity and severity of the flood. The study is carried out analyzing Twitter SMGI produced in June 2013 in Germany during the River Elbe's flood and applying statistical analysis in order to detect general spatial patterns in the occurrence of SMGI flood-related and associated with both its proximity and severity. The study demonstrates how SMGI contributed near the afflicted area shows a higher probability of being related to the event, providing reliable and useful information for managing disasters.

4.3.2 Political science, social science and media studies

Several studies concerning the political and social science domains proposing different approaches which benefit from SMGI for analysis, investigation of trends and prediction may be found in literature. In this domain Luce (2012), proposes a technical application to perform several analyses taking advantage of tweets. The event of French Presidential election in 2012 is analyzed accessing the Twitter stream API by the Pytolab application that performs several real-time contents analyses related to well-defined terms. The results expose both a report on textual analysis of relationships among keywords in data, such as parties and candidates, as well as several statistical reports about spatial and temporal patterns of the event. Moreover, the results exhibit how less than 1% of processed Twitter SMGI supplies geographical reference, leading toward a potential loss of information.

Relating the same domain, a study conducted by Tumasjan et al. (2010) analyzes Twitter information in order to evaluate the use of the social network as a forum for political deliberation and the representativeness of shared messages on the platform, namely tweets, as a valid mirror of users' political sentiment. The study concerns the analysis of over 100,000 tweets containing a reference to either a political party or a politician during the federal election of the national parliament in Germany on 2009. The approach proposed in the study utilizes textual analysis software implemented to assess emotional, cognitive, and structural components of text in order to profile political sentiment with regards to different dimensions, such as users' political orientation and emotions. The results show how the volume of tweets reflects the voters' preferences and may be considered similar to the traditional election polls results, while the sentiment of tweets is closely corresponding to political programs and candidates profiles. Despite the study dismisses the geographical components for the analysis and it presents limitations regarding the sample representativeness, the results demonstrate the capability for eliciting useful knowledge from social networks related to a specific application domain.

In the social science domain, Mandel et al. (2011) propose a study for analyzing the users' response to a natural disaster through Twitter. The authors collect over 65K of tweets related to the Irene's Hurricane during the period August 18th to August 31, 2011. The approach relies on the selection of on-topic tweets by applying keywords filtering technique and then by grouping the tweets by location and authors' gender. The aim of the study is to investigate different people responses to the disaster according to spatial distribution and demographic of users. In order to determine the tweets' location, the authors process the user location from tweet metadata, while to identify the users' gender, a comparison of users name with a list of the most used names in the United States is conducted. The findings show that the number of Twitter messages related to the Hurricane Irene exposes near real-time peaks in affected region, while the concern level in days after depends on the geographic area. Finally, a sentiment analysis has been carried out on the sample in order to identify major concern of females in respect to males with regards to the disaster event.

Along the same vein, a study proposing a semantic analysis of the Flickr tags, with the aim of determine whether the tag refers to a place or an event (Rattenbury et al., 2007) is found. The study takes advantage of statistical methods to allow the automatic identification of events or places from unstructured data, namely the tags, associated to photos by users. In spite of advanced analyses, the approach requires a strong manual intervention in order to extract and process Flickr SMGI.

Bertrand (2012) introduces a methodology to visualize the social graph from the Twitter contents, investigating the mutual relationships among contents and users on the social network. The approach relies both on the Twitter Stream API to perform an automatic extraction of data and on a graphing program to produce the social graph. The approach takes advantage of the mutual relationships among users in the sharing of tweets; however it dismisses the opportunity to investigate the geographic dimension of SMGI.

Furthermore, several investigate the user characteristics, behaviors and mutual relationships in order to identify potential patterns and communities in real-life and online social networks. In the case of online social networks, many studies analyze the Twitter users' behavior exploiting the homophily concept, which may be defined as the tendency of users to have ties with other people which are homogeneous regarding many socio-demographic, behavioral and intrapersonal characteristics (McPherson et al., 2001). A study proposed by Al Zamal et al. (2012) suggests a methodology to deduct a number of Twitter users characteristics by leveraging the homophily concept. The use of a machine-learning approach is proposed to investigate the characteristics of a users' sample via some well-noticed features of a classified sub-groups belonging to the sample. In practice, the study proposes a supervised machine learning approach by eliciting the users' characteristics, relying on the mutual relationships between classified and not classified users exclusively. The analysis is conducted on the social interactions of Twitter users, and demonstrates how the homophily phenomenon is strongly present also in the case of online community.

In a similar way, Kang and Lerman (2012) investigate the social interactions between Twitter users in order to analyze similarities in behaviors and discussed topics. The approach is implemented by means of a textual analysis on tweets' textual contents and demonstrates that a strong similarity insists between users having a follower-following or a mutual following relationship. Thus, the results demonstrate how the characteristics of a specific user may be inferred by the analysis of other users that have specific relationships with him.

In the domain of media studies, an application to explore in real-time the Twitter contents in search of events that attract users' concerns is discussed. The TwitInfo platform aims to identify media events according to well-defined terms of research, offering results in a timeline with an aggregate sentiment analysis of contents (Marcus et al., 2012). The data volume is processed in real-time, allowing the further exploration and investigation of detected events to users. Moreover, the application provides an analysis of spatial and sentiment patterns of the event, offering a glimpse of the users' opinions and the geographic area of interest for each specific event. In this case the approach leverages both advanced technology and analytical methods to deal with SMGI.

In addition, an application called TwitterMonitor to detect in real-time media trends from the Twitter stream is exposed. TwitterMonitor is developed in order to automatically detect high-rate keywords in tweets, namely bursty keywords, relying on the Twitter Stream API (Mathioudakis et Koudas, 2010). The analysis results are processed to extract further information about the event or the topic by means of advanced instrument, offering to users a potential summary of each detected event.

4.3.3 Urban and regional planning

In the urban and regional planning domain, a growing research body is concerned on the development of methodologies and techniques, which may elicit useful knowledge from SMGI for supporting design, analysis and decision-making in practice.

First, a study proposed by Frias-Martinez et al. (2012) aims to identify the land uses in urban environment relying of the Twitter users' dynamics. The approach is applied on the case study concerning the Manhattan urban area , relying on the Twitter Stream API to process the data volume in search of georeferenced tweets. In the methodology several spatial clustering analysis are performed by applying Self-Organizing Maps (SOM) (Oja and Kaski, 1999) and Voronoi Tessellation techniques (Voronoi, 1909), in order to obtain a partitioning of urban area according to the density of the SMGI contributions. Moreover, a set of temporal analysis is conducted on the SMGI timestamp in order to identify the potential land uses according to frequency and periods. Afterwards, a K-means clustering algorithm (Hartigan and Wong, 1979) is applied on the results of temporal analysis in order to assign a specific land use to each urban partition. Finally, a comparison between the study results and official data regarding the known land uses of the urban area is conducted. Results exhibit the capability of the approach to accurately detect commercial, residential and leisure areas, as well as other land uses, which are usually dismissed in official information, such as nightlife areas. The methodology fosters the use of advanced instruments and technology to automatically perform innovative analysis in urban planning, providing interesting results about the potential use of SMGI in this application domain.

Second, a study to investigate people movements and their preferences on landmarks in urban environment is reviewed. The approach aims to apply geovisual analytics techniques on Flickr photographs regarding urban landscapes and their related tags, in order to detect potential new city landmarks, or points of interest (POI), as well as the urban dynamics (Jankowski et al., 2010). The approach is applied on the case study of Seattle, relying on the Flickr API for performing the SMGI extraction and an exploratory spatio-temporal analysis about locations and time periods of the data. The procedural approach for processing the data involves three main steps. In the first step, several spatial clustering operations by means of Voronoi Tesselation are performed on SMGI to obtain maps of spatial distribution and the partitioning of the urban area, meanwhile the time periods of contributions are investigated on the basis of weekly intervals. In the second step, a set of statistical analysis are computed on SMGI dataset to identify potential interesting landmarks, according to several ranking criteria. Finally, the third step includes a set of analysis which aims to validate the obtained results. The obtained results argue that SMGI may be used as affordable source for the collection of geographic information related to urban area.

Third, Noulas et al. (2011) present a methodology to classify urban area and users' profiles by analyzing the Foursquare SMGI. In this case, the Foursquare SMGI is extracted through Twitter Stream API in order to overcome limits in data extraction caused by Foursquare API. Despite this approach may lead toward a loss of information, at the same time it stresses the tight relationships occurring between different social networks used by online community for different purposes. Foursquare is an online platform that allows users to check-in in specified places and locations and seamlessly share this information with their online social network over different social media platforms. In addition, each place or location available on the Foursquare service is classified in a specific category by the service itself, easing the detection of area presenting similar features. The aim of the study is the partition of London, United Kingdom and New York, United States urban areas and their consequent classification according to the semantic characteristics, namely the category of places, of Foursquare SMGI. The urban area partitioning is developed by means of a fishnet, which can be defined as a regular grid with imposed dimensions for cells, while the classification of each cell is based on the K-means algorithm, which leverages the Foursquare metadata associated to SMGI. Following the classification of urban area, the study proceeds with an users profiling analysis by means of spatial distribution and volume of shared contributions. The results allow the detection of similar areas in the same urban environment or in urban environment over different cities, as well as, the identification of different users' profiles.

Similarly, Cranshaw et al. (2012) discusses an approach to investigate and classify the urban area of Pittsburgh, Pennsylvania (United States) by means of Foursquare SMGI. The study is carried out applying a spatial clustering algorithm on Foursquare SMGI in order to identify areas which share common features in SMGI attributes. At the same time, the semantic component of Foursquare SMGI, namely the category of the places or locations, is used to assign each area to a specific group. In order to evaluate the methodological findings, the authors propose the assessment of urban areas classification with the results of semi-structured surveys conducted on a sample of locals. The assessment results show a high equivalence between results, showing that the classification obtained by SMGI is consistent with the perception of the inhabitants of the city. In addition, the study exposes how the official administrative divisions may not reflect the overall perception of the city by its inhabitants.

In another study concerning the urban environment investigation, Torres and Costa (2014) introduce a methodology to investigate and describe different activities conducted by users in public spaces of Rio de Janeiro (Brasil), thanks to the analysis of SMGI. The authors collect SMGI from different social platforms related to sport and leisure activities, such as Runtastic, MapMyRun, Runmeter, Smartrunner, Endomundo, Strava, Runkeeper, applying spatial, temporal and statistics analyses on the data to detect the different uses of the city's public spaces. The applied methodology is able to identify walking and bicycle path over the city, and at the same time it argues that the lack of suitable public spaces for sport and leisure may lead users to choose only specific streets and neighborhoods during specific periods of the day and week.

A methodology to develop a demographic analysis of London, United Kingdom by means of Twitter SMGI is proposed by Longley et al. (2014). The approach relies upon georeferenced SMGI extracted by means of the Twitter Stream API for the municipal area of great London. Several analyses are conducted on SMGI dataset in order to identify users' ethnicity and age, as well as their spatial distribution. The approach proposed by the authors leverages specific tools and software to conduct the textual analysis of SMGI contents, leading to the final classification of users, namely user profiling. In the study, 13 different variables related to both ethnicity (8 variables) and age (5 variables) are considered to perform the final profiling. Moreover, a temporal analysis is implemented in order to investigate the spatial patterns of the different identified users' groups, showing how different areas of the city may be differently experienced by diverse ethnic groups.

Along the same vein, a study conducted by Adnan et al., 2014 investigates the demographic characteristics of London (United Kingdom), Paris (France) and New York (United States) by means of Twitter SMGI. The approach proposed by the authors is similar to the one by Longley et al. (2014), but in this case the gender variable is also considered for the user profiling. The results of spatial, temporal and textual analysis show that the majority of classified users are male, while the dynamics and the spatial patterns in all considered cities are strongly affected by the users' ethnicity.

In another study, an investigation of the Instagram social network SMGI is proposed in order to analyze the potential differences in diffusion and patterns of use across several countries worldwide. Silva et al. (2013 A) analyzes the Instagram diffusion in different locations, such as New York (United States), Rio de Janeiro

(Brasil), Belo Horizonte (Brasil), Rome (Italy), Paris (France), Sydney (Australia), Tokyo (Japan) and Cairo (Egypt), detecting differences in spatial and temporal patterns. The findings of the study show variable patterns for different geographical area, as well as for different districts of a city, depending on cultural differences among users' habits. Moreover, a temporal analysis conducted on the Instagram SMGI allows the detection of specific patterns for each country and city, enabling the investigation of users' habits and daily routines, as well as the POIs discovery in the explored urban environments. The study concludes proposing the use of time analysis on social networks contents to identify behavioral and cultural differences among users of different countries and cities.

In a subsequent investigation, Silva et al. (2013 B) evaluated the differences between Foursquare SMGI and Instagram SMGI for depicting urban dynamics and user behaviors. This approach investigates the spatial distribution and the temporal patterns of SMGI collected for three main cities: New York (USA), San Paolo (Brasil) and Tokyo (Japan). The temporal patterns of contributions demonstrate to be strongly affected by the geographic area, but at the same time, even the social networks diffusion may be considered deeply influenced by this phenomenon. However, in the latter finding biases might be introduced by the collection method, which relies upon the API of a third social network, namely the Twitter Stream API, arising uncertainties in the validity of the assumption. Nevertheless, the results demonstrate how both Foursquare and Instagram may be suitable to identify the most popular areas in a city and to study the behavioral characteristics of users in different countries. On the one hand, Foursquare SMGI is more suitable to investigate the routes and the users' dynamics, while on the other hand, Instagram SMGI is more profitable to study the social dynamics within a city.

Finally, a study concerning the collection and analysis of SMGI by Campagna et al. (2013) introduces an ad-hoc developed platform, named Place I Care!, which may be considered as a VGI planning support tool for collecting information from concerned citizens about the environmental and socio-cultural spaces of a city, in order to ease participation and collaborative dialogue about urban issues. The authors develop a set of analysis with the aim of understanding what and where the main interests and concerns of the participants are, by leveraging the spatial, temporal and textual dimensions of SMGI. In addition, the approach proposes a novel analytics to be applied to the information stemming from the discussion among the participants, which includes: spatial and temporal analysis of user interests and concerns, spatial statistics on user preferences, multimedia content analysis, user behavioral analysis, or a combination of two or more of the methods, namely Spatio-Temporal-Textual Analysis (STTx). The novel analytics framework is developed in order to enable the elicitation of knowledge from people discussion in space and time. The study shows how the integration of SMGI with authoritative spatial data may be used to understand the local community's perception of the city, supporting discussion about the development of urban places, and potentially representing a way to inform official stakeholders, who lead decision-making processes.

4.4 Analytical opportunities for spatial planning

The findings provided by the number of reviewed approaches and studies found in literature, summarize the current development stage for advanced social media analysis methods and tools, which are easing the use of SMGI in several application domains. Focusing the attention on the spatial planning domain, several approaches and methodologies are proposed in order to take advantage of SMGI, which represents a potentially affordable and boundless source of information regarding people interests and concerns. Summarizing the findings of the reviewed approaches, it is evident that novel methodologies require advanced methods and tools for properly managing and analyzing the various facets of SMGI. As a matter of fact, SMGI exhibits a particular data structure, requiring the use of suitable tools and analytical methods to deal with the spatial, temporal and user dimension, as well as, to investigate the embedded multimedia contents. All the reviewed methodologies may manage and analyze the sheer volume of this data, although in several approaches a manual intervention is strongly required for guiding the analysis, the proper information extraction, or the results classification.

From an analytical perspective, the approaches introduced in the different domains are mainly based on the analysis of one or more dimensions of SMGI, namely spatial, temporal, textual or users, through the use of different methods and tools to elicit useful knowledge. Nonetheless, the approaches reviewed in the field of spatial planning, and in particular the one proposed by Campagna et al. (2013), stress the major opportunities arising from a parallel or integrated analysis of multiple dimensions, which may enable to inquiry more effectively the spatial and temporal patterns of contributions, as well as the social and urban dynamics, thus easing the investigation of users preferences and concerns in urban systems. Overall, the reviewed studies show how quantitative and qualitative analyses may be conducted on SMGI using spatio-temporal and statistical techniques to verify different hypothesis, unleashing the knowledge enclosed in the sheer volume of qualitative descriptive SMGI (Campagna et al., 2015). Indeed, the wealth of information available from social media about facts, opinions and feelings of users could affect the current practices in design, analysis and decision-making, and could inform smart strategies with a real-time monitoring of needs and requirements of local communities. In addition, the availability of geographic social network platforms may ease the processes of Public Participation, or Participatory GIS (PPGIS), both in technology and social terms. Recently, PPGIS initiatives have required major efforts in order to establish a suitable technological and management framework. By taking advantage of already available social networks and SMGI, both no technology setup and less commitment by the potential participants may be required, inasmuch involved participants voluntary use one or more social networks during their daily routines (Campagna et al., 2015). Nevertheless, it is important to underline how different combinations of analytical approaches may be required in order to interpret the local contexts proficiently thanks to the use of SMGI.

In the lights of these considerations, the summary of the results obtained by the reviewed studies and approaches, mainly in the domain of spatial planning, is provided in Table 1. In the table are exposed the different dimensions of SMGI and the main potential findings that an advanced analysis of each dimension, or multiple dimensions, may provide for eliciting knowledge in spatial planning.

SMGI DIMENSION	MAIN FINDINGS FOR SPATIAL PLANNING
Space	i) spatial patterns ii) partitioning of urban area iii) identification of POIs
Time	i) temporal patterns ii) temporal trends iii) temporal peaks
User	i) user statistics
Content (text, video, picture, etc.)	i) general interests and concerns ii) sentiment analysis
Space + Time	i) spatial-temporal patterns ii) identification of routes iii) urban dynamics iv) land-use classification
Space + User	i) user profiling
Space + Content	i) land-use classification
Time + User	i) user profiling
Time + Content	i) change in general interests and concerns
User + Content	i) user profiling ii) user interests and concerns
Space + Time + User	i) identification of user routes ii) identification of user POIs iii) geodemographics
Space + User + Content	i) user interests and concerns in space
Space + Time + Content	i) change in general interests and concerns in space
Time + User + Content	i) change in user interests and concerns
Space + Time + User + Content	i) change in user interests and concerns in space ii) (near) real-time information for spatial governance

Table 1. SMGI dimensions and opportunities for spatial planning analysis.

Moreover, the table summarizes the main results that advanced analyses on SMGI dimensions may provide for spatial planning in order to support analysis, design and decision-making. Despite SMGI may represent an innovative source of information regarding facts, opinions and preferences of users in space and time, it is important to be aware that this information may benefit from the official geographic information related to the referring context. Therefore, listed findings could be obtained by integrating SMGI with the available A-GI in order to extract profitable knowledge. In addition, the integration of datasets, originating from different social networks, might further improve the presented analytical opportunities.

4.5 Discussion

This chapter discusses the increased availability of SMGI over the global Internet and the opportunities that this type of information may disclose for analysis in different application domains, such as disaster and emergency management, political science, social science, media studies, as well as, urban and regional planning. In spite of the novel mechanism that VGI initiatives are fostering for the production and dissemination of geographic information, major concerns persist regarding the quality, reliability, accuracy and credibility of this information for practices and research. Indeed, VGI and its SMGI subset are heterogeneous data and usually unstructured data, which may enclose different knowledge basis, leading toward difficulties in the seamless use of this information for practices or in analytical framework. Nonetheless, VGI and SMGI may present important benefits in terms of affordability and timely data, as well as in the capability to provide information usually neglected in official information. Thus, several authors are concerned in the development of methodologies to evaluate and assess the quality and the fit-to-purpose of SMGI for different practices and research.

Furthermore, the reviewed studies found in literature were able to depict several methods and approach which may be conducted to investigate proficiently this type of information. SMGI requires advanced technologies, methodologies and ad-hoc tools to be analyzed and elicit knowledge; however, a lack of common methodologies or analytical frameworks to take advantage of this information for practices and research is detectable. In the spatial planning domain, the knowledge enclosed in SMGI may play a major role for supporting governance processes oriented to the local communities needs, fostering the development of smart cities initiatives tailored on real requirements of people and social dynamics. For this reason, the development of advanced tools, able to deal with the challenges of extraction, management and analysis of this information may be considered as the first milestone for easing and increasing the use of SMGI in practices. Despite of traditional data, SMGI refers to dynamic processes and requires new kind of tools to treat monitoring and decision-making in real-time about information that is continually changing, as well as, into finding suitable practices and procedures to integrate this experiential information with A-GI.

The reviewed studies provide several suggestions toward this direction and offer a glimpse concerning the current opportunities for a proper integration of SMGI with official information, as summarized above in Table 1. Altogether, the proposed methodologies and the different analytical opportunities contribute to show how SMGI might be used to elicit information, not only about the physical geography of places, but, overall, to investigate the perceptions of places and issues in time and spaces by the involved community, which may add a multifaceted perspective for spatial planning and decision-making.

CHAPTER 5

OBJECTIVES OF RESEARCH AND PROPOSED METHODOLOGY

5.1 Introduction

The increased availability and production of SMGI over the global Internet are paving the way to innovative analysis scenarios in urban and regional planning, as depicted by the findings of the several approaches found in literature. Operationally, the integration of SMGI with A-GI may allow the development of analyses based upon quantitative and qualitative information, enriching the current capabilities of traditional spatial planning analytical methodologies. Commonly, urban and regional planning processes need large amounts of information for developing sustainable decision-making and implementing public policies. This information may be official or derived from direct observations or questionnaires conducted on a representative sample of population. Nevertheless, the traditional methods for gathering information may be highly expensive and time consuming, limiting the capability to have frequently updated information (Frias-Martinez et al., 2012; Jankowski et al., 2010). Hence, the data commonly used to study social phenomena is static and reports the particular instant at which the information is collected, dismissing opportunities to provide change in interests or preferences over a period of time (Antony, 2010).

SMGI may represent an innovative way to deal with the requirements for updated datasets on urban environments (Goodchild, 2007) or to favor the collection of opinions and requirements from local communities (Williams, 2010), easing social participative practices (Miller, 2006). However, the 'Big Data' nature of SMGI may require ad-hoc tools and analytical methodologies to take advantage of the enclosed knowledge. As a matter of fact, one major issue to extract useful knowledge from these innovative sources is to find an efficient way to manage the avalanche of information. The management issues for information from Big Data sources (Caverlee, 2010) is giving rise to an emerging new research field, namely Computational Social Science (CSS) (Lazer et al., 2009). This discipline may be described as the "integrated and interdisciplinary pursuit of social inquiry with emphasis on information processing and through the medium of advanced computation" (Cioffi-Revilla, 2010). Therefore, the discipline exploits a computational approach to the social sciences, that is the use of advanced instruments, tools and models to enable the collection and the analysis of massive amounts of data (Lazer et al., 2009). CSS is starting to revolutionize the way research is carried out, affecting both the empirical work by means of 'big data' and the theoretical model through computer simulation models for investigating social phenomena (Hilbert, 2015). The main CSS domains are automated information extraction systems, social network analysis, social GIS, complexity modeling, and social simulation models. Nonetheless, significant barriers persist to the advancement of

CSS, but an increased availability of user-friendly tools and methodologies might magnify the extent of CSS to other domains of interest (Lazer et al., 2009).

5.2 Research objectives

The research objectives concern the SMGI Analytics formalization, in order to exploit the massive wealth of SMGI for spatial planning analyses and governance, and the development of advanced instruments for easing the SMGI Analytics application in practices. In fact, one of the major hurdles limiting the SMGI use in practices is the lack of easy-to-use tools, able to extract, to manage and to integrate this information with A-GI, allowing the development of contextual and multi-dimensional analyses. In the light of the above considerations, the methodological approach is developed following two main directions:

(1) the design and development of ad-hoc tools able to deal with the issues regarding the access, management and analysis of SMGI;
(2) the formalization of the novel SMGI Analytics framework to support design, analysis and decision-making in spatial planning.

The two research objectives are strictly intertwined, inasmuch the development of tools is concerned to the implementation of the SMGI Analytics framework in practices. Therefore, the first step of the research deals with the development of a user-friendly tool, called SPATEXT, which is able to extract information from multiple Social Media, while providing several functionalities to implement spatial, temporal and textual analysis, as well as, to take into account the user dimension.

5.3 Tools for SMGI Analytics: SPATEXT

The opportunities for the use of SMGI in spatial planning guide the design of a user-friendly suite of tools, called SPATEXT (SPAtial-Temporal-tEXtual Toolbox), which eases the extraction and management of information from multiple social media platforms and the contextual integration in a GIS environment for analysis. The SPATEXT suite is implemented as Python 2.7 add-in for the commercial software ESRI ArcGIS©, including a number of ad-hoc developed tools, which may be used to (1) retrieve SMGI from social networks (including Twitter, YouTube, Wikimapia, Instagram, Instagram Places, Foursquare and Panoramio); (2) geocode or georeference data; and carry out analyses on the (3) spatial, (4) temporal, (5) textual and (6) user dimension of SMGI. In addition, the analytical methods available in the tool include several clustering algorithms in order 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 A-GI, 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 hurdles regarding access, management and analysis of 'big data' and may be consequently categorized in three different classes:

1) data collection;

2) data management;

3) data analysis.

The first class includes user-friendly tools that enable information harvesting 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. 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. An overview of the SPATEXT functionalities is presented in Table 2, where the main tools are classified and briefly described according to the specific class functionality, while the SPATEXT architecture is shown in Figure 4.

SPATEXT SUITE TOOLS				
CATEGORY 'DATA COLLECTION'				
		query parameters		
Tool Name	**Function**	**space**	**time**	**keyword**
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	✓		

CATEGORY 'DATA MANAGEMENT'			
		function activation	
Tool Name	**Function**	**Manual**	**Automatic**
Geocode address	Geocoding place/address from string	✓	
Geocode table	Batch-Geocoding place/address from table		✓
Georeferencing	Georeferencing SMGI coordinates		✓
Decomposition tools	Decompose SMGI shapefile in multiple shapefiles	✓	✓
Google™ Static Maps	Add Google Static Map URL in SMGI attribute fields	✓	

CATEGORY 'DATA ANALYSIS'	
SPATIAL AND CLUSTERING ANALYSIS	
Tool Name	**Function**
DB-SCAN (Density-Based SCAN)	Run DB-SCAN algorithm (Ester et al., 1996) on SMGI feature class to detect density clusters and add the cluster group in a new field of SMGI feature class
Feature-Based DB-SCAN	Run FB-DBSCAN algorithm (Ester et al., 1996) on SMGI feature class to detect density clusters for each group in SMGI feature class (e.g. User)
Advanced SQL maximum	SQL selection on a SMGI feature class to detect maximum values for a specific group (e.g. User Cluster exposing max number of points)
TEXTUAL ANALYSIS	
Tool Name	**Function**
Attribute to string	Creation of a text file from an attribute field in a SMGI feature class for textual analysis
Attribute to table	Creation of a table from an attribute field in a SMGI feature class for textual analysis
Attribute to tag-cloud	Tag-clouding analysis from an attribute field in a SMGI feature class
Selection to tag-cloud	Tag-clouding analysis from a spatial/attribute selection in a SMGI feature class
Text to tag-cloud	Tag-clouding analysis from a text file
TEMPORAL ANALYSIS	
Tool Name	**Function**
Identify Month/Weekday/Day/hour	Add the Month/Weekday/Day/hour of creation in a new field of SMGI feature class
Trend Day/hour	Creation of a 24h/60min time graph and statistic report from SMGI feature class

Table 2. SPATEXT tools.

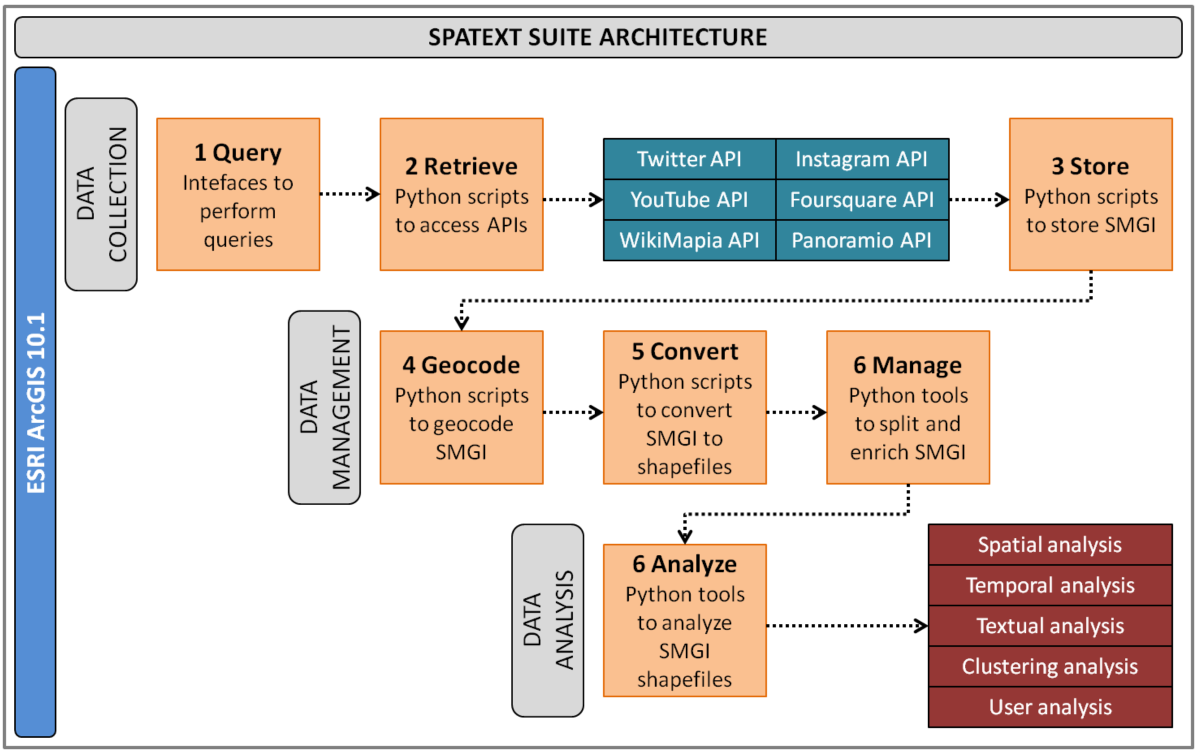


Figure 4. SPATEXT architecture design.

The tools included in SPATEXT are designed and developed in order to supply user-friendly tools to planners easing the development of spatial planning analyses concerning the use of SMGI for practices. Planning and especially urban planning are generally based on technical sectors and share some methods and tools with social sciences (Goldstein and Garmin, 2006). The discipline concerns both policies and practices, including a scientific basis built upon sociology, economics, environmental, sciences, geography and GIS (Zanon, 2014). Any advance in these sectors requires a specific expertise that further delineates a planner's toolbox (Zanon, ibidem).

Nevertheless, the SMGI nature and the methods and tools required for exploit this information in practice, usually overtake the traditional competencies of a planner, which should deal with issues concerning programming, database creation, management and query on 'big data'. In this respect, the SPATEXT suite is developed to address these later issues supplying a set of tools that might be proficiently used in practices to solve several problems concerning the access and the management of SMGI, and to ease the development of analytical processes.

5.4 Proposed methodology: shaping SMGI Analytics

Following the development of the SPATEXT suite, the second research objective is the development of an adequate analytical framework to take advantage of SMGI in practices, namely SMGI Analytics. The SMGI Analytics relies upon the integration of A-GI from the Regional SDI and SMGI from different social networks in order to investigate spatial, temporal and user dimensions of contributions. At the same time, the analytical framework aims to exploit the textual contents, embedded in SMGI, for eliciting opinions, preferences and requirements of local communities.

The SMGI Analytics builds upon an investigation method that includes descriptive spatial analysis and exploratory SMGI analyses, encompassing STTx (Campagna, 2014 A), in order to proficiently elicit knowledge from SMGI for spatial planning analyses and decision-making. In addition, the analytical methodology proposes the use of clustering and geodemographic segmentation techniques to develop further analysis able to investigate users' preferences and dynamics into urban environment from the SMGI spatial, temporal and user dimensions. Operationally, the SMGI Analytics uses SPATEXT suite tools to deal with the issues of SMGI collection, management and analysis, as well as to integrate this information into GIS environment.

The SMGI Analytics framework consists of the following stages, which shape the operational workflow to carry out multi-dimensional analysis on SMGI for spatial planning:

1. **data collection**: SMGI is extracted from several social networks by SPATEXT tools directly in a GIS environment for a specific geographic location, temporal period or natural language query. Extracted SMGI is seamlessly georeferenced, managed, converted and stored in shapefile format to enable GIS analyses. Each social network provides a specific structure for SMGI;
2. **exploratory spatial-temporal analyses and A-GI integration**: the SMGI spatial and temporal components are investigated directly in GIS environment, in order to explore potential spatial and temporal patterns of interest in the area and the local community dynamics. In this stage, the SMGI dataset is integrated with several official datasets from the Regional SDI of Sardinia to guide further analyses;
3. **textual analyses**: a set of textual analyses are conducted on the SMGI dataset in order to elicit further information that may provide answers to detected spatial and temporal patterns, as well as, to gain insights on users' perception, opinions and requirements;
4. **spatial-temporal-user cluster analyses**: several clustering methods are carried on the SMGI dataset in order to identify, classify and interpret high density clusters of SMGI contributions. For this purpose, the Density-based Spatial Clustering of Applications with Noise (DB-SCAN) algorithm (Ester et al., 1996) and a slightly modified version called Feature-based DB-SCAN (FB-DBSCAN) are conducted on the SMGI dataset. This way, the resulting clusters may be further investigated inquiring the local community or the individual interest toward certain spaces in diverse time periods;
5. **user profiling**: a geodemographic segmentation is applied on the geographical area taking advantage of official census statistics (ISTAT). The approach is based upon the K-Means algorithm (MacQuenn, 1967; Hartigan and Wong, 1979) in order to group census tracks according to several key characteristics, which are common to the resident population. The segmentation's results are then coupled with the cluster analyses results in order to assign a specific label to each user contributing to the SMGI dataset. The final result is a user profiling based on geodemographic variables, which allow specific analyses regarding concerns and preferences of different population groups.
6. **complementary SMGI extraction**: the SMGI dataset is integrated with further social networks data for the same location or time period, in order to explain the reasons behind a number of detected clusters or specific users groups' dynamics.
7. **explanation of phenomena**: the results of traditional GIS analyses coupled with SMGI analyses are used to explain users' preference, opinions and requirements in the investigated area of interest. Findings may inform decision-making process and territorial policies, coupling official and experiential knowledge.

The aforementioned SMGI Analytics stages allow the investigation of spatial and temporal patterns, the identification of Points of Interest (POI), the elicitation of users preferences and opinions, the investigation of land uses and urban dynamics, the user profiling, as well as the inquiring of changes in users interests and concerns in space and time. Despite the major opportunities for spatial planning analyses, the main innovations proposed in the approach are the integrated multi-dimensional analyses of SMGI, as well as, the analysis of the user component, namely the user profiling. The latter kind of analysis may rather disclose increased opportunities for the development of social phenomena investigation than traditional analytical methods, allowing the study of specific population segments behaviors and preferences.

The disclosed analytical opportunities may represent a valid instrument to support analysis, design and decision-making in geodesign studies, as well as, to inform 'smart city' strategies with near real-time information about the concerns of users. The linearity along the several stage of the SMGI Analytics is not fundamental and different analyses may be conducted with a number of feedbacks or shortcuts according to a study's requirements. The SMGI Analytics framework logic is shown in figure 5.

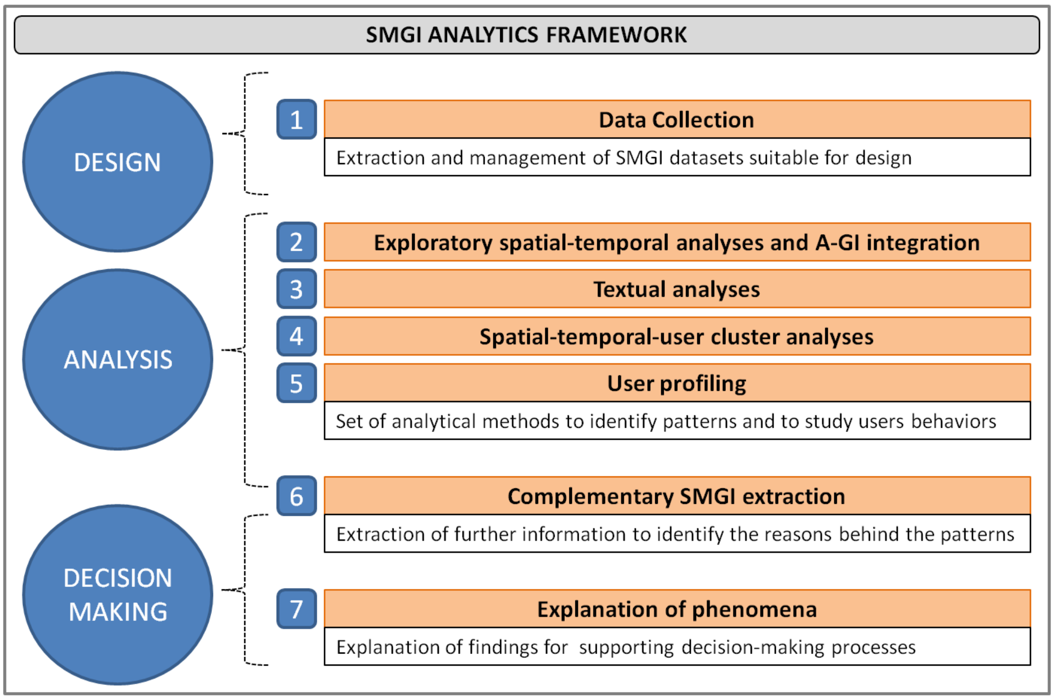


Figure 5. SMGI Analytics framework logical structure.

5.5 Data collection

Data collection is executed by means of the several social networks extraction tools provided by SPATEXT, which enable the SMGI harvesting through the spatial, temporal or natural language queries. The tools provided in SPATEXT allow the SMGI extraction from a number of major social networks such as: Twitter, YouTube, Instagram, Instagram Places, WikiMapia, Foursquare and Panoramio. Effectively, any SMGI is characterized by a specific data structure, derived from the database structure, the Application Programming Interface (API) architecture and, the societal commercial and legal frameworks. Commonly, the publicly available SMGI data model features only a subset of the original attributes or multimedia data produced by the user, implying that the analytical potential is in general greater within the social media companies than for the public (Lazer et al. 2009; Campagna, forthcoming). Therefore, each tool is designed and implemented in order to carry out suitable queries for exploiting most of the available knowledge according to the aforementioned barriers to data extraction.

The extracted data are automatically converted in a dataset suitable for further investigation in GIS environment. As a matter of fact, the tools parse the extraction results and seamlessly produce a points' dataset wherein the geographic coordinates of each SMGI are coupled with the available attributes and links to multimedia data. The general SMGI dataset structure, produced by SPATEXT tools for each social network, is provided in Table 3.

Social Network	**Attributes**	**Type**	**SMGI Dimension**
Twitter	Tweet geotag Tweet text Tweet data User registration User location Country Place full name	Lat/Long String Data Data String String String	Space Multimedia content Time Time User Space Space
YouTube	Video geotag Title Author Description Data production Video URL	Lat/Long String String String Data String	Space Multimedia content User Multimedia content Time Multimedia content
Instagram	Photo geotag Place name User name User ID User Picture Photo URL Photo data Photo UNIX timestamp Number of Comments Number of Likes Photo Caption	Lat/Long String String Integer String String Data Integer Integer Integer String	Space Space User User User Multimedia content Time Time Multimedia content Multimedia content Multimedia content
Instagram Places	Place geotag Place name Place ID	Lat/Long String Integer	Space Multimedia content Multimedia content
Wikimapia	Place geotag Place Title Place ID Location Name Location Municipality Place URL Description	Lat/Long String Integer String String String String	Space Multimedia content Multimedia content Space Space Multimedia content Multimedia content

Foursquare	Place geotag Place name Place ID Location city Location country Location code Place typology Place short description Number of check-in Number of users	Lat/Long String Integer String String Integer String String Integer Integer	Space Multimedia content Multimedia content Space Space Space Multimedia content Multimedia content Multimedia content Multimedia content
Panoramio	Photo geotag Photo title Photo upload Author Author ID Photo URL Photo ID	Lat/Long String Data String Integer String Integer	Space Multimedia content Time User User Multimedia content Multimedia content

Table 3. SMGI dataset structures.

For each social network, SPATEXT produces a SMGI dataset including different attributes related to a specific multimedia content by means of the different APIs and limitation. It should be noted that, while certain attributes express quantitative measurements, other provide qualitative information related to the multimedia content. Therefore, the selection of a specific social network for the extraction should be based upon previous awareness of available attributes, which represents the intent of the social platform and the goal of the analysis.

5.6 Explorative spatial-temporal analyses and A-GI integration

The second SMGI Analytics stage concerns the development of explorative analyses in order to investigate spatial and temporal patterns in the SMGI dataset, as well as, the local community dynamics. At this stage, the SMGI dataset should be integrated with several official datasets from the Regional SDI concerning the purpose of the study.

A set of simple investigation on the spatial distribution are conducted in order to evaluate differences in density of contribution and to identify the locations attracting the major interest from the users. This type of explorative spatial analyses is usually carried out by means of traditional GIS instruments and is required in order to develop further punctual investigations on SMGI contents.

At the same time, the temporal component is investigated for different time periods, searching for potential peaks of interest, trends and dissimilarities in the use of the social network by the users. The explorative temporal analyses are carried on inquiring SMGI contributions, aggregated for seasons, months, days of the week and hours of the day, in order to disclose interesting patterns, eventually.

5.7 Textual analyses

Commonly, the use of textual analytics techniques enable the visualization of relevant posts on a map (Berry and Kogan, 2010) the discovering of underlying knowledge, when the amount of information grows rapidly, such as in the case of SMGI. The SMGI Analytics third stage introduces two simple textual analytic methods, namely tag-cloud analysis, or text-cloud, and semantic analysis. The tag-cloud is the visualization of word frequency in a weighted list and graphical form, and it is suitable to depict the most used words, namely tags, in a text, recognizing possible hidden information. This visualization technique, originated in the 21st century (Manovich, 2010), is popular to visualize prominent topics in presentations, political speeches and texts (Lamantia, 2008). Generally, the result of tag-cloud analysis is a representation of a weighted word list in alphabetical order, where the importance of each tag is shown through font size or color. (Kaser and Lemire, 2007; Halvey and Keane, 2007). In order to enable the tag-cloud analysis, a set of SPATEXT tools allows the investigation of the most used words in SMGI textual contents, delivering a graphical results and a statistical report.

Semantic analysis is the process of coupling syntactic structures from the levels of the writing to their language meanings. In the case of SMGI, a semantic analysis is conducted in order to the identification of an underlying set of common categories in the results of tag-cloud analysis. As a matter of fact, the ability to extract semantics can improve current tagging systems allowing more powerful search (Rattenbury et al., 2007), as well as, helping with knowledge elicitation. Despite a complete semantic understanding of tags is unlikely, the opportunity to assign some structure to SMGI textual contents, may help to investigate users' preferences, opinions and requirements. The semantic analysis is manually conducted on the tag-cloud results in order to elicit further information.

The aim of the textual analysis is to identify words, which can be considered directly related to toponyms, activities or feelings, in order to obtain insights on users perception, as well as, for allowing further analysis on detected topics. Tag-cloud analyses coupled with semantic analyses are used in order to obtain information about the users' perception at the regional and the local scale. The textual analyses searched for the most used words both on SMGI textual contents, investigating dissimilarities in order to improve the understanding of detected words. The resulting tag clouds show different words for the considered textual contents, but the semantic analysis leads the identification of an underlying set of common categories. Several words can be considered as 'noise' caused by different languages and sentence structures, however, most of the words may belong to four main categories:

(1) toponyms/places;

(2) activities;

(3) values;

(4) links/URL.

In the first category the words referring to specific toponyms (city name) or physical places (urban location) are considered, while in the second category the words related to events or objects are grouped. The third category encloses the words related to adjectives used for personal evaluation of places and events, and in last category the words relative to external links, persons, blog or URL are enclosed. The insights obtained through textual analyses allow the development of further specific analysis, concerning people spatial and thematic perception of places.

5.8 Spatial-temporal-user cluster analyses

The results of explorative spatial and temporal investigations, as well as, of textual analyses, may guide the development of further analyses for inquiring the geography and the urban dynamics of the study area. Especially the major density of SMGI in certain areas fosters the development of analytical methods to identify, classify and interpret the users' interest toward these specific spaces. For this purpose, SMGI Analytics introduces a set of clustering analyses, which may deal with the requirements of interpretation for some aggregated SMGI contributions in specific geographical areas. Clustering analysis is a generic name assigned to a variety of mathematical methods, which may be used to identify which objects in a set are similar (Romesburg, 2004). These mathematical methods accomplish a cluster analysis by sorting real objects processing the data description of certain attributes. The DB-SCAN algorithm (Ester et al., 1996) and a slightly modified designed version called Feature-based DB-SCAN (FB-DBSCAN) are integrated in SPATEXT, for accomplishing the clustering analysis on a SMGI dataset and for enabling the SMGI clustering based on the spatial density of points.

5.8.1 Clustering analysis

Since the late 1970's and early 1980's a growing research body is concerned with the development of clustering analysis methods to classify real objects in different groups (Romesburg, 2004). Clustering is a procedure to place individual resources into groups on the basis of quantitative information concerning one or more characteristics that are inherent between different resources (Van der Walt and Barnard, 2006). This kind of procedure is useful when it is necessary to process a large quantity of data in order to successful interpret the similarities among the analyzed original objects (Hart, 1982). An important distinction should be made between clustering analysis and clustering methods, or clustering algorithms. As a matter of fact, a clustering method is exclusively the used methodology for the partitioning of items in clusters, whereas a cluster analysis refers to a wider sequence of steps to follow for completing the analysis (Milligan, 1996).

One of the main issues in clustering analysis is the choice of a suitable method, or clustering algorithm, to perform the analysis, inasmuch there is no method that may be considered more appropriate than another one for the purposes of a specific study (Openshaw and Gillard, 1978). Operatively, the most appropriate clustering algorithm for a specific problem should be selected experimentally, unless a mathematical reason eases the choice (Estivill-Castro, 2002). For a proper clustering method's choice it should be necessary to consider a set of factors such as: (1) the aim of the clustering procedure; (2) the comparison of results between different clustering methods (Grekousis and Thomas, 2012); and (3) the evaluation of clustering results by means of performance criteria (Michie and al., 1994). In fact, a clustering algorithm that is designed for a specific data model, may suffer when dealing with completely different data sets leading toward inappropriate results (Openshaw and Gillard, 1978).

The real notion of cluster is difficult to be defined due to a number of different cluster models, which require the use of specific clustering algorithms (Estivill-Castro, 2002). However, the main cluster models used for the development of clustering analysis in literature are:

- connectivity models: are based on the principle that items are more related to nearby items than to farther items. The model relies upon hierarchical clustering algorithms (Romesburg, 2004; Johnson, 1967), which provide an extensive hierarchy of clusters that may merge each other according to a specific distance, while the results are commonly figured as a dendrogram. The algorithms process data on the basis of different way to compute distances, such as: minimum distance, maximum distance, mean distance, to name a few;
- centroid models: represent clusters by means of central vectors, which may also not belong to the original data set. When the number of clusters is fixed to a k value, k-means clustering algorithms (Forgy, 1965; MacQuenn, 1967; Hartigan and Wong, 1979) are used in order to optimize the resolution problem. K-means clustering aims to partition objects into k clusters, wherein each object is assigned to the cluster with the nearest mean. Unfortunately, most of the centroid based algorithms, such as the K-means, require specifying in advance the number of clusters k.
- distribution models: define clusters by assigning objects that belong most likely to the same statistical distribution. Distribution based clustering may capture correlation and dependence between data attributes, but in real data it may be burdensome to define or identify a mathematical model. The expectation–maximization (EM) algorithm (Dempster et al., 1977) is commonly used to model clusters by means of a fixed number of Gaussian distribution to fit the data characteristics.
- density models: define clusters as areas of higher density than the remainder of the data set. One of the most popular density based clustering method is the Density-Based Spectral Clustering of Application with Noises (DB-SCAN) (Ester et al., 1996). Similarly to the connectivity models, the

algorithm connects points within certain threshold distances, but only if density criterion, defined as the minimum number of other objects within a specified radius, is contemporary satisfy by them. The algorithms provide the same results in each run, limiting the number of required iterations, while the objects in sparse areas are considered as noise, commonly. Nevertheless, these algorithms suffer in detecting intrinsic cluster structures which may exist in real life data.

Despite the analytical opportunities arising from the cluster models and the associate algorithms, a cluster analysis should be conducted taking into account several steps, which involve multiple decisions by the analyst, in order to obtain a successful result (Milligan and Cooper, 1987). Moreover, each possible decision leads toward an alternative result that may be more or less suitable for the purpose of the study (Lorr, 1983). Therefore, each cluster analysis should record and report any taken decision and the reasons behind (Vickers and Rees, 2007). An example of methodological framework for cluster analysis is proposed by Milligan (1996) and later simplified by Everitt et al. (2001). The methodological framework consists of seven different steps:

1) clustering elements: are the objects to cluster that should be representative, possibly covering the whole geographic area of study;
2) clustering variables: are the measurements taken on certain objects characteristics to take into account during the analysis. Variables should be informative and able to discriminate among objects;
3) variables standardization: is required only for certain data sets if the clustering variables present particular characteristics;
4) measure of association: represents a measure of similarity or dissimilarity to be selected in respect of the clustering variables, reflecting the degree of proximity or separation between objects;
5) clustering method: is the clustering algorithm to apply on the dataset and should be robust and able to detect the clusters suspected to be present in the real objects;
6) number of clusters: is the number of different groups to find out in the data and it represents the most difficult decision of the analysis (Miller, 1996; Vickers and Rees, 2007). Different rules may help the analyst in decision; however the choice is usually based upon experience, usefulness, as well as, clustering method.
7) interpretation, testing and replication: is an assessment of whether the solution meets the needs of the investigation. In addition, in this stage test for sensitivity and replication of the analysis may be conducted to confirm the results.

The SMGI Analytics proposes a workflow based upon the aforementioned steps to analyze and investigate the spatial distribution of the SMGI dataset detecting clusters by means of contributions' density. The approach relies upon the clustering density models using the DB-SCAN algorithm as clustering method for analysis, in order to identify which areas attracting the major interest of the users in the urban environment. These areas may be considered belonging to clusters of high contributions' density. In addition, a clustering analysis on the SMGI contributions of each single user is conducted relying on spatial and temporal components, leading toward the identification of unmapped buildings into A-GI and of residential areas.

The methodological approaches proposed in this paragraph are further described in the following chapters by means of several example case studies.

5.8.2 The Density-Based Spatial Clustering of Applications with Noise (DB-SCAN)

The Density-Based Spatial Clustering of Applications with Noise (DB-SCAN) (Ester et al., 1996) is a density based clustering algorithm, which takes advantage of the density-reachability concept to partitions points in groups. The DB-SCAN algorithm offers major advantages with respect to other clustering algorithms. First, it is not necessary to know a priori the number of clusters, which also may differ in size and shape. Second, DB-SCAN works using exclusively two parameters: ε (eps) that is the maximum threshold distance for including points in the same cluster, and min_pts (minimum number of points), which is the minimum count of nearby points to define a cluster.

The density-reachability concept behind the algorithm states that if a point *q* is directly reachable from a point *p* within a given threshold distance, namely ε (eps), and at the same time *p* is surrounded by a sufficient number of points (min_pts), then *p* and *q* may be considered objects of a cluster. The algorithm processes the dataset starting from a random point and evaluating the number of other points within the ε-neighborhood with the min_pts value. If the number of points is larger than min_pts value, a new cluster is defined otherwise the analyzed point is labeled as noise.

Then the analysis continues with another not visited random point in the dataset. However, the algorithm processes each point of the dataset multiple times, raising the opportunity for a noise-labeled point to be later found in a sufficient ε-neighborhood of another point, becoming part of a cluster.

The following provided pseudocode broadly describes the logic underlying the DB-SCAN algorithm. The algorithm is described using a simplified fictional programming language in order to ease the readability and comprehension of the analytical steps.

```
DBSCAN(Dataset, eps, minPts) {
  C = 0
  for each P in Dataset {
    if P is visited
      continue next point
    mark P as visited
    NeighbPts = regionQuery(P, eps)
    if sizeof(NeighbPts) < minPts
      mark P as NOISE
    else {
      C = next cluster
      expandCluster(P, NeighbPts, C, eps, minPts)
    }
  }
}
```

```
expandCluster(P, NeighbPts, C, eps, minPts) {
  add P to C
  for each P' in NeighbPts {
    if P' is not visited {
      mark P' as visited
      NeighbPts' = regionQuery(P', eps)
      if sizeof(NeighbPts') >= minPts
        NeighbPts = NeighbPts joined with NeighbPts'
    }
    if P' is not in any cluster
      add P' to C
  }
}
regionQuery(P, eps)
  return all P within P's eps-neighborhood (including P)
```

In spite of notable advantages, the DB-SCAN algorithm presents a set of disadvantages. The algorithm is not entirely deterministic for border points and its results depend upon the distance measure used for calculations. Usually, the Euclidean distance is used as common distance metric, but for high-dimensional dataset it may be not fully suitable (Beyer et al, 1999). Notwithstanding, the SMGI Analytics framework introduces the use of this algorithm in order to detect high-density clusters among the SMGI dataset, allowing the discovery of urban areas attracting the major interest of local communities, the identification and classification of individuals area of interest, the detection of unmapped buildings in official datasets, as well as, the development of a method to potentially identify users' location.

5.9 User profiling

Nowadays, the generation of user profiles from samples of users' interests and inherent characteristics is a common task for Artificial Intelligence (AI) research (Krulwich, 1997). In many applications, ranging from recommendation systems to web marketing, the building of individuals personalized profiles may represent an important opportunity for different tasks. Indeed, the user profiles may be used to recommend music, movies, or other products, to retrieve documents according to user interests, as well as, to guide marketing campaigns and targeted advertisement (Krulwich, 1997; Li et al., 2012).

The SMGI Analytics framework implements a user profiling methodology in order to assign a specific group label to each user contributing to the SMGI dataset, actually enabling analyses concerning the interests and preferences of different population groups. The user profiling methodology concerns the coupling of results obtained from the spatial-temporal user clustering analyses, introduced in the fourth stage of SMGI Analytics, with the results of an ad-hoc implemented geodemographic classification, or geodemographics, based upon the official census demographic information for the Sardinia region (Italy). The cluster analyses results may provide useful information to identify the users' residential location, while the geodemographic classification enables the population partition in groups according to socio-demographic characteristics, available from official statistics for each census track. Coupling the results of the two methodologies may disclose opportunities to assign each user contributing to the SMGI dataset to a specific population group, enabling further advanced analyses on diverse groups' preferences. In addition, the proposed geodemographics may have a wide variety of applications, ranging from geographic analysis to social marketing and consumer profiling.

5.9.1 Geodemographic classification

Geodemographic classifications are small area classifications which provide summary indicators of social, economic and demographic characteristics, while partitioning examined areas into groups on the basis of similarities among their common inherent features (Vickers and Rees, 2007; Adnan et al., 2010). The term was coined in the 1970's referring to novel public sector approaches for identifying deprived city areas by means of classification techniques, which bring together spatial patterns from a range of variables (Webber and Craig, 1978). During the 1980's, geodemographic approaches gained popularity in the private sector, enabling the development of target marketing methodologies (Birkin et al., 2002). Nowadays, geodemographic classifications, or geodemographics, may be defined as "the analysis of people by where they live" (Sleight, 1997; Webber, 2004). This kind of technique builds on the principle that population and place are inextricably linked. Therefore, knowing where an individual lives may reveal profitable information about that person (Vicker and Rees, 2007). In addition, relying upon the assumption that similar people may more likely live in similar places, knowing information about an individual may potentially enable to infer information concerning other individuals in the same place (Webber, 2004; Weiss, 2000). In the light of these considerations, a geodemographic classification creation may be considered an intense computational activity based upon the use of clustering methods for partitioning large multidimensional datasets into different groups exposing a high inter-group similarity among dimensions (Adnan et al., 2010).

Commonly, geodemographic classifications partition neighborhoods at two or more hierarchical level relying upon census data and potentially any secondary data sources (Harris et al., 2005; Adnan et al., 2010). Most of the commercial geodemographic classification systems exploit census data, financial data, property information and behavioral or attitudinal surveys; however, other geodemographics builds exclusively on official census data. Famous geodemographic classification systems are: CAMEO UK, ACORN, MOSAIC, geoSmart system, the Tapestry Segmentation and the Output Area Classification.

The CAMEO UK system is a geodemographic classification system, developed and maintained by Callcredit Information Group Ltd, that proposes a number of consumer classifications for the United Kingdom. The

CAMEO UK segments consumers into 10 key groups at the first level and 68 distinct categories at the second level (Callcredit Information Group Ltd, 2015).

The ACORN classification system, developed by CACI Ltd, segments the United Kingdom population in 5 general categories, which are further subdivided into 17 groups and finally in 56 types. The systems exploit more than 400 variables originated from government and consumer data (CACI, 2014).

MOSAIC is a worldwide geodemographic classification system which partitions the population in a three-tier classification composed by 15 groups at the first level, 67 household types at the second level and 155 person types at the last level (Experian, 2010).

The geoSmart system is a geodemographic classification developed by RDA Research for segmenting Australian population relying on socio-economic status variables according to a two-tier classification. The first level is composed by 7 groups, while the second level consists of 54 segments (RDA Research, 2010).

The Tapestry segmentation systems takes advantage of census and consumers' data to partition the United States neighborhoods into 67 distinct market segments related to socio-economic and demographic variables. In addition, the segments are grouped into 14 groups, which finally define 6 urbanization groups (ESRI, 2015).

Finally, the Output Area Classification (OAC) (Vickers and Rees, 2007) is a free and open geodemographic system for segmenting the UK's population. The system uses exclusively official census data to identify 41 different variables, which are used to develop a three-tier classification. The first level consists of 7 groups, the second level presents 21 groups, meanwhile the third level contains 52 groups.

Despite the notable number of available geodemographic classification systems, partly for the intensive computational activities and partly for the required commercial confidentiality, the creation of such systems is commonly limited to expert professional. As a matter of fact, most commercial systems use closed methods, avoiding to document in details the methodological steps, the clustering methods and the provenience of input data (Harris et al., 2005; Adnan et al. ibidem; Vickers and Rees, 2007; Singleton and Longley, 2009). Therefore, major opportunities for research arise from the OAC system. OAC is characterized by procedural openness enabling the base for a scientific reproducibility as demonstrated by the studies proposed by Singleton and Longley (2009) and by Petersen et al. (2011) that are both based on the OAC methodology. Different clustering methods may be applied in developing a geodemographic classifications, but no algorithm may ensure to achieve the best result and overall no theoretical certainty about their real classification efficiency is available (Grekousis and Thomas, 2012). As a matter of fact, clustering techniques may rely on neural networks, genetic algorithms, fuzzy logic, as well as on K-means algorithms (Forgy, 1965; MacQuenn, 1967; Hartigan and Wong, 1979), which represent the most applied methods.

On the basis of these considerations, the thesis proposes a geodemographic classification of the Sardinian region (Italy), developing a methodological approach similar to the one proposed by Vicker and Rees (2007) for implementing the OAC system.

5.9.2 The K-Means algorithm

The K-Means algorithm is a method of vector quantization that partitions the n observations of the input dataset into an established k number of clusters, by assigning each observation to the cluster with the nearest mean. The algorithm aims to partition the n observations, represented as a d-dimensional vector, by minimizing the within-cluster sum of squares, as described by the following formula:

$$\arg\min_S \sum_{i=1}^{k} \sum_{x \in S} \|x - \mu_i\|^2 \qquad \mu_i = mean\ of\ points\ in\ S_i$$

The algorithm is based on an iterative refinement technique that alternates two steps: (1) assignment step, or expectation step, and (2) update step or maximization step (MacKay, 2003). During the first step each observation is assigned to the cluster presenting the least within-cluster sum of squares, namely the squared Euclidean distance or nearest mean. During the second step, a new mean is calculated for the centroid of observations belonging to each cluster. The K-Means achieve the final result when no observation is further assigned to a cluster or after an established number of iterations. Despite the procedure is able to achieve a final result, the K-Means algorithm does not necessarily find the most optimal configuration. This phenomenon may depend by the algorithm's significant sensitivity to the initial random selection of cluster centroids. In order to deal with this issue, the K-Means++ algorithm (Arthur and Vassilvitskii, 2007) avoids this random choice by exploiting the distance matrix between observations in order to detect the most appropriate centroids in the input dataset. The K-Means++ algorithm is used in the research as clustering method for the geodemographic classification.

5.9.3 SMGI Analytics Geodemographics

The aim of the SMGI Analytics Geodemographics is the partition of the Sardinian census tracts relying on a number of socio-economic variables derived from the official census data. Afterwards, the resulting classification is used in order to allow the SMGI dataset users profiling. The geodemographic classification is conducted using exclusively the official census data provided by the Italian National Institute of Statistics (ISTAT) for the year 2001. Despite the existence of more recent information hailing from the 2011 national census, the geographic coverage of this information is not suitable for the purposes of the study. Secondary sources of non-census data are dismissed from the analysis in order to avoid potential issues concerning spatial scales, accuracy and reliability.

According to the methodology proposed for the development of OAC classification by Vickers and Rees (2007) in the UK, the first stage of the analysis concerns the identification of significant variables, able to discriminate among the socio-economic characteristics of the population. By reviewing the available census data at the census tract geographic scale, 43 variables are identified and assigned to 5 main categories:

- demographics;
- household composition;
- housing typology;
- socio-economic status;
- employment condition.

Despite the official census data offers 199 different variables, the final list presents a significant variables reduction due to a detailed assessment of each variable according to a number of principles (Vickers et al., 2005). For a proper selection it is fundamental to evaluate: 1) presence of high correlation among variables; 2) existence of badly behaved distributions; 3) opportunity to merge variables in composite variables; 4) the constancy of variables across the whole Region; 5) the vagueness or uncertainty in measurements; 6) the uninteresting spatial distribution; 7) the temporal consistency; 8) the opportunity to standardize variables affected by age distribution, as well as, 9) the opportunity to standardize all variables.

In the light of this assessment, Table 4 lists the selected variables for computing the geodemographic classification.

VARIABLE	DESCRIPTION	CATEGORY
V1	Percentage of resident population aged 0-4 years	Demographics
V2	Percentage of resident population aged 5-14 years	Demographics
V3	Percentage of resident population aged 25-44 years	Demographics
V4	Percentage of resident population aged 45-64 years	Demographics
V5	Percentage of resident population aged > 65 years	Demographics
V6	Percentage of resident population Afrikans	Demographics
V7	Percentage of resident population Asian	Demographics
V8	Percentage of resident population Caucasic of Hispanic	Demographics
V9	Density Pop/sqKm [normalized by range method]	Demographics
V10	Percentage of unmarried	Household composition
V11	Percentage of married	Household composition
V12	Percentage of separated, divorced or widowed	Household composition
V13	Percentage of single person household	Household composition
V14	Percentage of couple household	Household composition
V15	Percentage of household of 3-4 persons	Household composition
V16	Percentage of household > 5 persons	Household composition
V17	Percentage of house owners	Housing Typology
V18	Percentage of house tenants	Housing Typology
V19	Percentage of house with heating	Housing Typology
V20	Average rooms per house [normalized by range method]	Housing Typology
V21	Average persons per room [normalized by range method]	Housing Typology
V22	Atypical houses [normalized by range method]	Housing Typology

V23	Average area per house [normalized by range method]	Housing Typology
V24	Percentage of old houses 1919 - 1971	Housing Typology
V25	Percentage of recent houses > 1972	Housing Typology
V26	Percentage of building with 1 or 2 roofs	Housing Typology
V27	Percentage of condominium	Housing Typology
V28	Percentage of building with > 3 apartment numbers	Housing Typology
V29	Percentage of people with High education level	Socio economic
V30	Percentage of people with Medium education level	Socio economic
V31	Percentage of people with Low or No education level	Socio economic
V32	Percentage of commuters	Socio economic
V33	Percentage of inner-municipality commuters	Socio economic
V34	Percentage of employed	Socio economic
V35	Percentage of Students	Employment condition
V36	Percentage of Not employed (retired or other condition)	Employment condition
V37	Percentage of businessman of freelancer	Employment condition
V38	Percentage of salaried worker	Employment condition
V39	Percentage of employers in Agriculture	Employment condition
V40	Percentage of employers in Industry	Employment condition
V41	Percentage of employers in Public Services	Employment condition
V42	Percentage of employers in Trade, Restaurant, Transport, Communication	Employment condition
V43	Percentage of employers in Financial intermediation and business	Employment condition

Table 4. SMGI Analytics Geodemographics: selected socio-economic variables.

The geodemographics has the purpose to define a two-tier hierarchical classification of the Sardinian population in order to enable a synthetic description of the different population groups. The first hierarchical level is able to provide a general description of groups in the Region, while the second hierarchical level aims to highlight the peculiarities of each sub-group in more detail.

The methodological approach builds on multiple iterations of a K-Means algorithm taking advantage of the *grouping analysis* tool of ESRI ArcGIS. The *grouping analysis* relies on the K-Means++ algorithm (Arthur and Vassilvitskii, 2007) in order to optimize the analysis initiation avoiding the random pick of centroids in the input dataset. This capability enables the algorithm to converge more properly to a suitable result, that is to achieve a final maximized inter-cluster distance and minimized intra-cluster distance. In addition, the *grouping analysis* tool may enable the evaluation of further variables related to spatial distribution and spatial auto-correlation among observations (Assunção et al., 2006); however this capability is dismissed in order to avoid potential issues in classification arising from hidden spatial patterns. Depending on the established number of clusters, the tool identifies the most appropriate observations to be considered as cluster centroids for the following iterations. Each observation, namely the census track, is assigned to the cluster presenting the least within-cluster sum of squares, namely the squared Euclidean distance of the 41 selected variables. After the first assignment of all observations, the tool calculates the new cluster centroids and starts a new iteration. The tool runs for a set number of iteration or until the convergence achievement that occurs when no observation is reassigned to another cluster.

The main problem lies in identifying the correct number of clusters for partitioning the input dataset. In fact, it is not possible to define in advance the proper k clusters to segment the observations. However, according to the methodological approach found in literature, the analysis is performed iteratively by varying the number of clusters for each run of the tool. Several authors suggest about 6 groups for the first hierarchical level and 20 sub-groups for the second level (Callingham, 2003 in Vicker and Rees, 2007). Hence, the grouping analysis is tested for a k number of clusters ranging from 4 to 8 for the first hierarchical level. The results of each run are evaluated considering two factors:

1) the pseudo F statistic, namely the ratio between-cluster variance to within cluster variance (Calinski and Harabasz, 1974);
2) the cluster size, that is the number of observations belonging to each cluster for different k.

The pseudo F statistic is calculated using the following formula:

$$Pseudo\ F = \frac{(GSS)/(k-1)}{(WSS)/(N-k)}$$

where:
K = number of clusters
N = number of observations
GSS = between-group sum of squares
WSS = within group sum of squares

The evaluation of cluster size play an important role in geographic classification, inasmuch it is preferable to have clusters as closely sized as possible to each other. Indeed, unevenly sized clusters, overall in the first hierarchical level, may be problematic when clusters are further segmented for creating the next hierarchical level. The factors evaluation suggests how the segmentation for k = 6 shows the more appropriate results in terms of both Pseudo F statistic and number of observations belonging to each cluster when dealing with the Sardinian geodemographic classification. Once the first hierarchical level is defined, the same procedure is used to further partition each identified group. In order to obtain about 20 groups for the second hierarchical level, the grouping analysis is tested for k ranging from 2 to 5 and the results are evaluated in respect to the two aforementioned factors. Finally, the analysis of the second hierarchical level leads toward the identification of 18 sub-groups for the whole Sardinia Region.

After the final geodemographic classification, each group and sub-group is analyzed to find a suitable name or description able to highlight the inherent peculiarities. In order to identify a significant name, the variables of each group are evaluated with the mean regional values, highlighting the main similarities and dissimilarities. An example of sub-group 1 comparison is provided in Figure 6, where a histogram quantifies the major differences between the sub-group variables and the regional mean. In addition, the summary table 5 lists the variables that expose the most positive or negative differences from the Regional variable means.

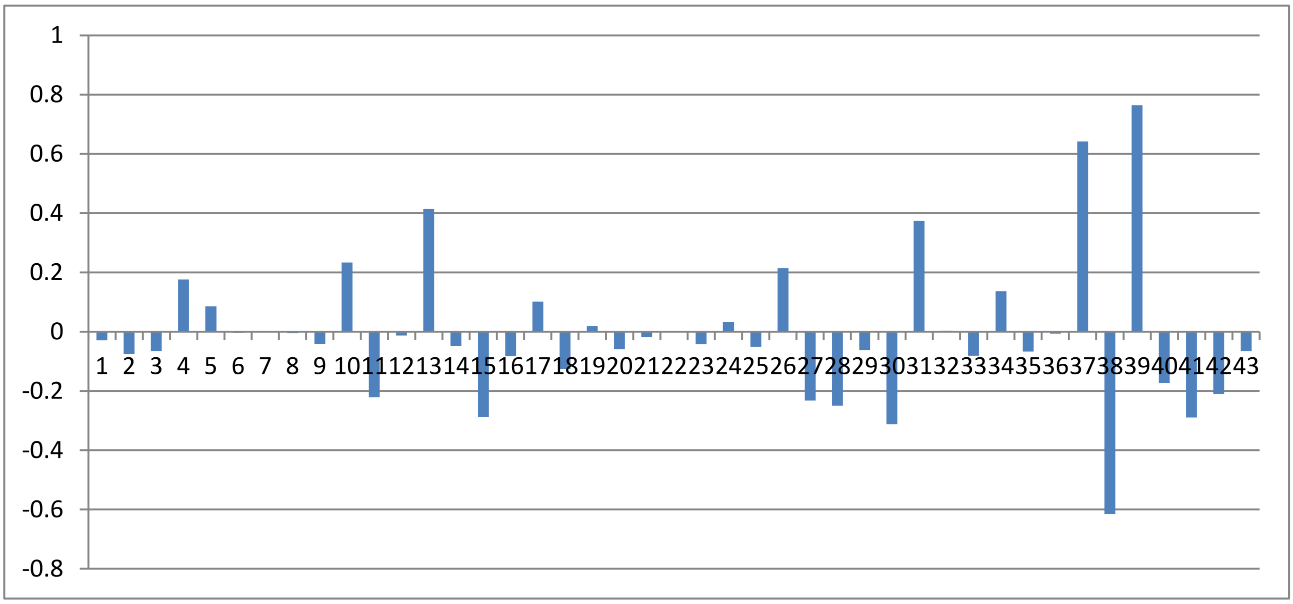


Figure 6. SMGI Analytics Geodemographics: sub-group 1 in Sardinia.

V10	Unmarried	+	Household composition
V11	Married	-	Household composition
V13	Single person household	+	Household composition
V15	Household of 3-4 persons	-	Household composition
V26	Percentage of building with 1 or 2 roofs	+	Housing Typology
V27	Percentage of condominium	-	Housing Typology
V28	Percentage of building with > 3 apartment numbers	-	Housing Typology
V31	Percentage of people with Low or No education level	+	Socio economic
V37	Percentage of businessman of freelancer	+	Employment condition
V38	Percentage of salaried worker	-	Employment condition
V39	Percentage of employers in Agricolture	+	Employment condition
V41	Percentage of employers in Public Services	-	Employment condition
V42	Percentage of employers in Trade, Restaurant, Transport, Communication	-	Employment condition

Table 5. Geodemographics: variables exposing main differences from mean in sub-group 1.

In addition, a short natural language label is assigned for each sub-group in order to highlight the main socio-economic characteristics of the involved population. However, it is difficult to set significant names to the sub-groups profiles, which cannot capture more than a tiny part of the clusters distinctiveness.

After the Sardinia geodemographic classification is concluded, the obtained findings are coupled with the results of the spatial-temporal user cluster analyses in order to obtain the user profiling. Therefore, relying upon the clustering results concerning the users' dwelling areas locations, each user is assigned to his specific geodemographic group. In the case of users not resident in Sardinia, the label is 'Italian' if the provenance is in Italy or 'Tourist' if the user is resident worldwide.

The final user profiling is then used to investigate spatial or temporal dynamics among different population groups, such as:

- In which period a specific group is more likely to visit a specific geographic location?
- Which are the behavioral similarities between users in the same group?
- Which areas attract the major attention of a specific population group?
- At which time certain areas are visited?
- Which areas accommodate the preferences of multiple groups?

The investigation of these questions may disclose innovative opportunities for spatial planning and governance, supporting the development of further analyses aiming at understand local communities' preferences and foster the development of 'smart strategies' oriented to assess specific locations.

5.10 Complementary SMGI extraction

The complementary SMGI extraction in the SMGI Analytics methodology aims to extract further knowledge from social networks about interesting areas or time periods, in order to ease the explanation of specific phenomena, as well as to answer the aforementioned questions. Coupling information from multiple sources may allow to identify specific Points of Interest (POIs) in the study area or may ease the evaluation of differences in preference. Usually this stage is performed at the local scale for answering specific questions about local dynamics or patterns. In the thesis, this step is performed with the purpose of finding answers to particular spatial patterns of population groups by extracting information from two location-based social networks, namely Instagram Places and Foursquare.

5.11 Explanation of phenomena

The last step of the methodology is the explanation of identified phenomena. Relying upon the integration of official information and SMGI it is possible to obtain insights that traditional spatial planning methodologies are unable to find or that may require high expenses or time consumption. Indeed, SMGI is able to provide updated and near real-time information able to depict the current dynamics affecting the area of study.

5.12 Discussion

This chapter discusses the methodological approach adopted in the thesis, introducing an innovative methodological framework to analyze A-GI and SMGI to support design, analysis and decision-making for spatial planning and governance. From an operational perspective, the integration of SMGI with A-GI may enable analyses based on quantitative and qualitative information, offering opportunities to gain insights about opinions, preferences and requirements of local communities.

The goals of the methodological approach are to evaluate at different geographic scales the opportunities arising from SMGI extracted from different social network to foster the detection of spatial and temporal patterns and local community dynamics, enriching the traditional spatial planning methodologies with novel analytical methods. In order to achieve these goals the methodological approach consists of 7 steps, ranging from the initial data collection to the final explanation of detected phenomena.

The methodology deals with several hurdles limiting the use of SMGI in practices, such as the issues regarding data extraction, data management and data analysis. In order to address these issues, the methodology follows two main directions. The first followed direction concerns the development of a user-friendly tool called SPATEXT which may offer a solution for dealing with the extraction and management problems, as well as, for providing novel analytical tools. The second direction introduces the novel SMGI Analytics framework that may guide to a proficient SMGI use in practices.

The SMGI Analytics framework proposes a set of analyses, which exploit traditional GIS spatial tools and the novel tools introduced by SPATEXT for developing spatial analyses at different geographic scales. In addition, in order to deal with the issue of data mining from sources as SMGI, which presents an inherent 'big data' nature, the methodological approach takes inspiration from CSS techniques to develop several clustering analyses for identifying areas of general interest, areas of individual interest, urban dynamics, as well as, to detect potential user residential locations.

Moreover, the introduced SMGI Analytics proposes the use of a geodemographic classification in order to develop a user profiling classification, which may enable further analyses for answering specific questions useful for spatial governance and 'smart cities' strategies at the regional and local scales. The proposed method follows the guidelines used in the study of Vickers and Rees (2007) for creating the Office for National Statistics 2001 output area classification in the UK.

In literature, several studies propose novel approaches for analyzing SMGI, but no one introduces a complete methodological framework that aims to enable spatial planning practices to exploit the integrated use of SMGI and A-GI from the data extraction toward the final decision-making processes, relying upon multi-dimensional analyses, which consider spatial, temporal and user dimensions of the geographic information. In the next chapter, several SMGI Analytics applications at different geographic scales are proposed on a number of case studies in order to demonstrate the capabilities and the opportunities of this novel methodological approach.

CHAPTER 6

SMGI ANALYTICS FRAMEWORK: EARLY EXPERIENCES

6.1 Introduction

The wealth of information available from social networks about users' movements, opinions and preferences may affect several domains of interest where the subjectivity of observation is relevant for expressing the views, the needs, the call of individuals and communities. The increasing SMGI production and availability over the web may foster innovative analysis scenarios in spatial planning and geodesign, which in turn could be used to increase smart city performances.

The SMGI Analytics framework is introduced as a novel analytical methodology to foster the use of SMGI in spatial planning domain in order to support design, analysis and decision-making processes. The SMGI Analytics prepares the ground for the integration of A-GI and SMGI enabling the development of multi-dimensional analyses able to elicit insights and knowledge from the experiential information of users, which traditional spatial planning methodologies may struggle at obtaining. Nonetheless, the assumptions and the novel analyses introduced by the SMGI Analytics framework should be tested in order to assess the real opportunities and capabilities of the approach. Therefore, the thesis discusses the SMGI Analytics application on several case studies, conducted at the global, regional and local scale, relying upon a number of the described methodological stages for obtaining useful insights and knowledge from the SMGI, as well as, for testing the functionalities of the SPATEXT instrument that eases to carry out the methodological workflow.

The SMGI Analytics application by SPATEXT is discussed through a number of example case studies conducted at different geographic scales, investigating both local communities' perceptions on relevant topics for spatial planning and the geography of places. The application of the SMGI Analytics for SMGI processing is discussed regarding:

1- the analysis of the Cleopatra cyclone in Sardinia (Italy) through Twitter SMGI at the global scale investigating the spatial and temporal patterns of social network contributions;
2- the investigation of local community perceptions on the Sardinian landscape at the regional scale by analyzing YouTube SMGI in order to evaluate different perception of landscape by users in different geographic zones;
3- the analysis of users' perception on several neighborhoods of the Cagliari municipality in Sardinia (Italy) through YouTube SMGI and relying upon spatial and textual analyses at the local scale;

4- the investigation of the dynamics and preferences of users in the urban environment of the Iglesias municipality, Sardinia (Italy) through Instagram SMGI at the local scale. In addition, the case study introduces a clustering methodology in order to identify the areas in the urban environment attracting the major users' interest, as well as, to detect the residential users' locations.

The different case studies are carried out in order to achieve different goals. First, the case studies are developed in order to test the SPATEXT tools capabilities for extracting, managing and analyzing SMGI from different sources. Second, the case studies exploit SMGI extracted from various social networks, in order to assess the SMGI Analytics capabilities to elicit useful insights from different sources, as well as, to establish the more appropriate scale for analysis depending upon the chosen social platform. Third, different novel clustering approaches for identifying areas and POIs are introduced and evaluated in practices, disclosing opportunities to investigate the replicating opportunities and the fit-to-purpose of this method. Finally, the findings of the case studies are used to properly formalize the SMGI Analytics framework that is then used in a more complex case study during the following of the research.

6.2 Example case study Twitter: the Cleopatra cyclone

The first experience with SMGI Analytics concerns the illustrative case study of the cyclone Cleopatra in Sardinia (Italy) through Twitter SMGI. Collected SMGI may enclose perceptions, opinions and needs from the social network users during the occurrence of the cyclone in November 2013 that deeply affected several areas of Sardinia. The integration of A-GI with the personal perspectives provided by SMGI may offer meaningful information for investigating the spatial and temporal patterns of contributions, as well as, to eventually support design and decision-making. The example case study is conducted at the global scale inquiring differences in textual contributions provided by users from different countries. In addition, the temporal trend is analyzed in search for both eventual peaks of interest and the reasons behind these peaks. Hence, the results of the data collection, the spatial-temporal analysis and the textual analysis on SMGI are proposed, assessing the main features of the SPATEXT suite and the opportunities arising from the SMGI Analytics.

6.2.1 Data collection

The dataset about the cyclone Cleopatra event is collected by the "Tweet Extractor" tool of SPATEXT, setting the keywords "Sardinia floods" in English language to avoid disturbance for harvesting because of too generic keywords. The extraction is performed at the global scale and grants the extraction of 399 different tweets, approximately 16 and 24 hours after the start of the event (2013.11.19 from 8:30 to 9:30 +0000 UTC and from 17:30 to 18:00 +0000 UTC). The tool automatically stores the public tweets and their metadata in a data table including text, user name, user location, user registration data, tweets creation time, and if available, country, place name, as well as, the geographic coordinates. In general a tweet may have three kind of geographic references: i) the spatial tag collected by the GPS mobile sensor; ii) the user location given in her/his profile; iii) any toponym given in the text. Each type of location implies different level of accuracy, with regards to the semantic of the relationship with the content, and should be treated accordingly.

Unfortunately, only 0.01% of extracted information contains geographic coordinates, thus SPATEXT geocoding tools are used to address this issue, by processing the user location metadata. The percentage of georeferenced tweets is smaller than 0.77% that commonly distinguishes the percentage of tweets with geographic component (Semiocast, 2012), but this may be due to time and language constraints. The geocoding approach for SMGI may introduce positional uncertainty shifting the geographic information from tweet to user, but at the same time it may provide the spatial distribution of the topic, according to the users' spatial distribution.

6.2.2 Spatial-Temporal analyses

The SPATEXT geocoding tool is used to process the tweets data table in order to generate a point feature dataset containing the locations of the 399 collected tweets, allowing the opportunity to develop further spatial analysis by integrating SMGI with other spatial data layers. A check on the dataset points out how 42.6% of collected tweets expose wrong or inexistent places for user location, demonstrating the reluctance of certain users for disseminating personal information over the web.

The spatial analysis is conducted studying the spatial distribution of the topic among worldwide countries with the aim of discovering dissimilarities and interesting spatial patterns. The most active countries in the discussion (in English) about Cyclone Cleopatra are the United States, the United Kingdom, Italy, Nigeria, and India, with tweets percentage of 22.5%,6.5%, 6.0%, 2.5%, and 1.7%, respectively. These values may be considered representative of the different degree of interest of users about the topic among these countries. An assessment of this hypothesis may be addressed through the analysis of the different percentages. The USA percentage exhibits the higher value (> 300%) than the UK, Italy, Nigeria and India, but this peak of interest may be explained considering that the USA are ranked 1st in Twitter top countries classification with over 140M accounts at the time of the study. Similarly, the percentage of the UK could be explained considering the 4th position in Twitter rank with over 30M accounts. On the other hand, the Italian percentage (6.0%) raises interesting questions about the spatial influence of the topic. In fact, the percentage is both affected by language constraint in the data collection and by the less importance of Italy in Twitter top countries (about 5M accounts). The ration between the percentage of on topic contributions and the number of Twitter Italian accounts exposes the 1.2 value that is the highest among the considered countries, exposing a notable level of interest for the topic by Italian users. Nonetheless, this notable data volume concerning the cyclone Cleopatra in Italy may be easily explained considering the affected geographical areas, namely the Sardinia that is one the twenty Italian regions.

Moreover, the percentages of Nigeria and India raise further questions related to the spatial distribution of the topic on Twitter. In fact, both countries are affected by language constraints and do not expose any direct spatial relationships with the topic. Hence, these values are further investigated through temporal and textual analyses in order to understand potential hidden dynamics and relationships.

The analysis on the spatial distribution of dataset is conducted through standard spatial analysis tools in GIS environment. The dataset used as input for the analysis is directly developed by SPATEXT, demonstrating how the tool may be suitable for a seamlessly extraction and integration of SMGI in GIS environment. In Figure 7 the spatial distribution worldwide of the collected tweets about the topic is provided. Several tweets are depicted with different colors and symbols due to their belonging to the different examined countries. The spatial distribution may be useful to gain further insights and understand potential hidden dynamics, guiding the development of analyses on the temporal distribution of SMGI, that are hereafter discussed.

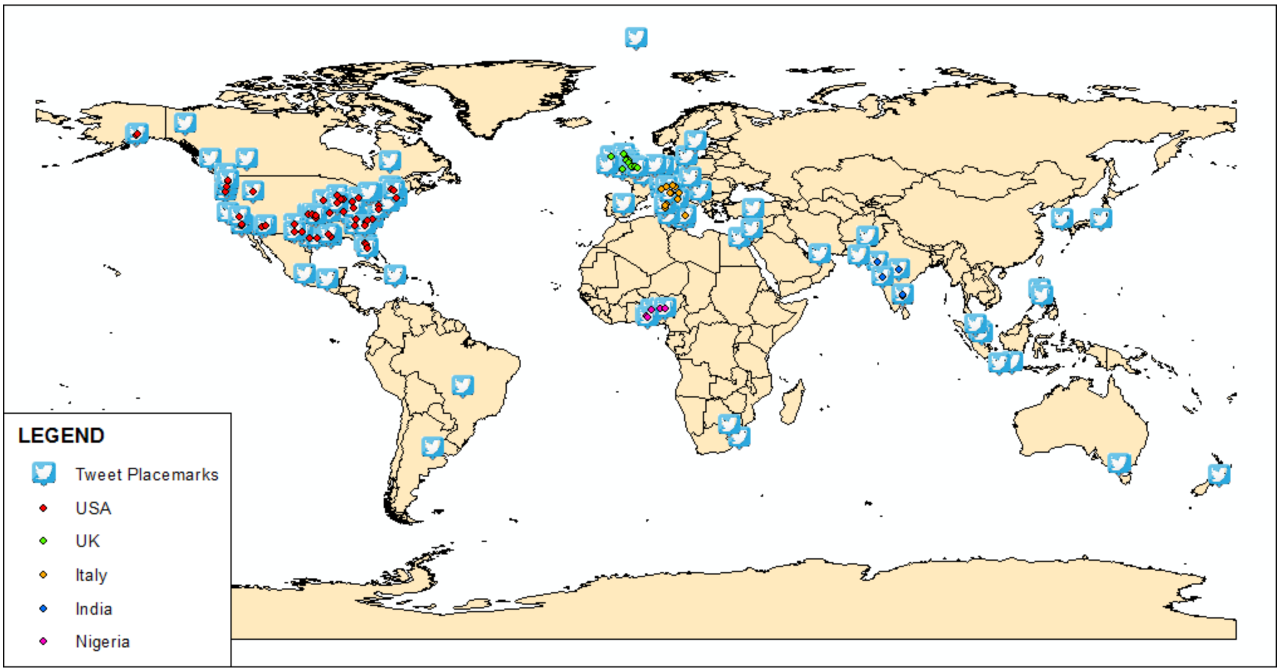


Figure 7. SMGI Spatial distribution: analysis conducted on the most representative countries.

The introduction of temporal analyses relies on the use of the tweets' metadata on creation time in order to discover meaningful patterns and peaks of interest about the topic. The temporal component is investigated in search of meaningful patterns about the topic according to two different approaches: (1) investigation of the spatial distribution for the first 30 tweets and (2) analysis of temporal trend by 5 minutes periods.

First of all, the SMGI temporal distribution is processed in order to evaluate a potential direct relationship between the creation time and the geographic areas. The spatial distribution of the first 30 created tweets is analyzed, but the results point out a heterogeneous spatial distribution, so dismissing the existence of any meaningful spatial pattern or direct relationship between space and time in the SMGI dataset.

Secondly, the other approach analyzes the temporal distribution of the SMGI dataset aggregated for the two extraction periods (from 8:30 to 9:30 +0000 UTC and from 17:30 to 18:00 +0000 UTC November 2013), that defines 17 periods of 5 minutes. The analysis' results on the extraction periods show the presence of 199 tweets in the first period and 200 tweets in the second one, suggesting an increasing interest about the topic along time.

At the same time, the analysis on the 17 periods is computed in order to identify potential anomalies in the temporal distribution and to investigate the temporal trend through a time-series graph. The results show peaks of interest in the 9th and 13th period that are further investigated for validation. The validation of a peak of interest is based on the contemporary fulfillment of two criteria in order to avoid false positives. The first criterion for the validation of a peak requires a value for the time period >=100% than the specific value of linear regression in the same period, meanwhile the second criterion requires a value for the period =>100% than the average value. Both the identified peaks satisfy the criteria; therefore they are validated as peaks of interest in the temporal trend and further investigated by means of textual analyses. The results of SMGI temporal analysis is provided in Figure 8, showing the time-series and the detected peaks of interest.

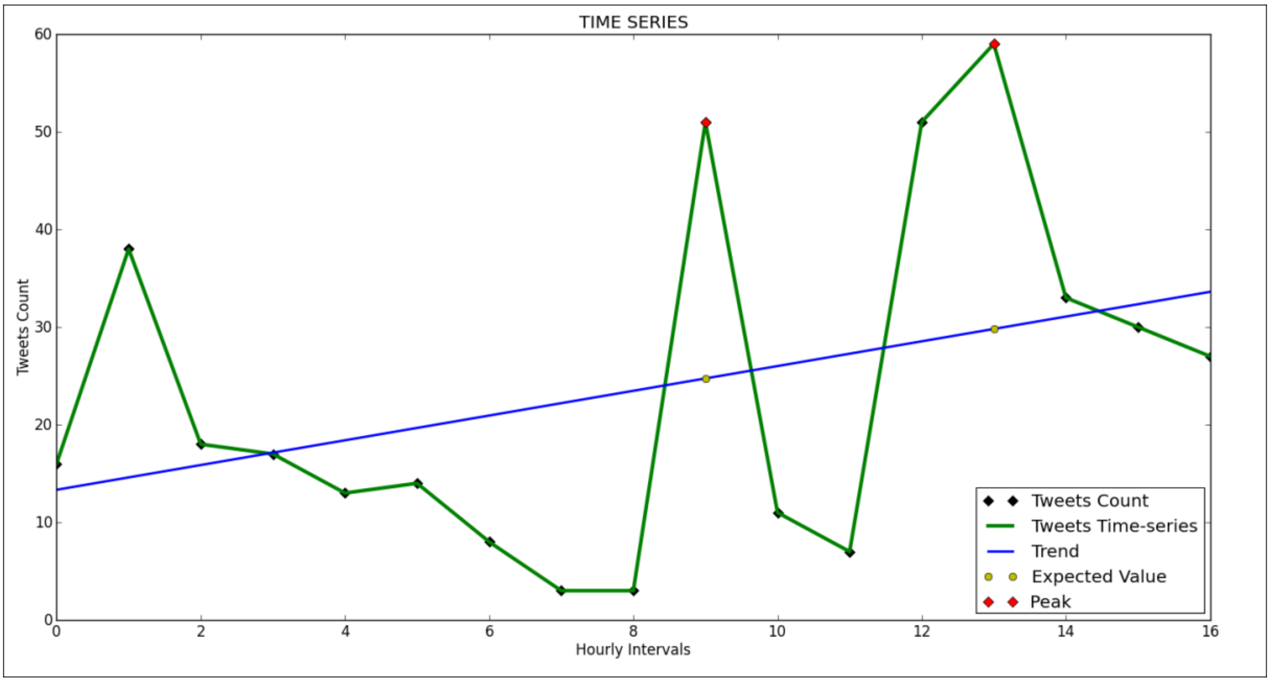


Figure 8. SMGI Temporal distribution: temporal trend and peaks of interest.

6.2.3 Textual analysis

The peaks of interest, detected by the temporal analysis, oriented the analysis toward the study of further aspects on the SMGI dataset. A textual analysis is conducted on the dataset to discover further knowledge in order to find potential answers to peaks. The tag-cloud analysis tool of SPATEXT is used to discover the most used words in the SMGI dataset. The tag-cloud is the visualization of word frequency in a weighted list and graphical form, thus it is suitable to depict the most used words in a text for recognizing possible hidden information. The textual analysis is conducted both on worldwide tweets and on tweets from Italy, investigating possible differences in contents. Furthermore, the SMGI contributions related to the Nigeria are investigated in order to discover the potential causes behind the spread of the topic in this country. The resulting tag-cloud visualizations are provided in Figure 9, for Italian and worldwide tweets, respectively.

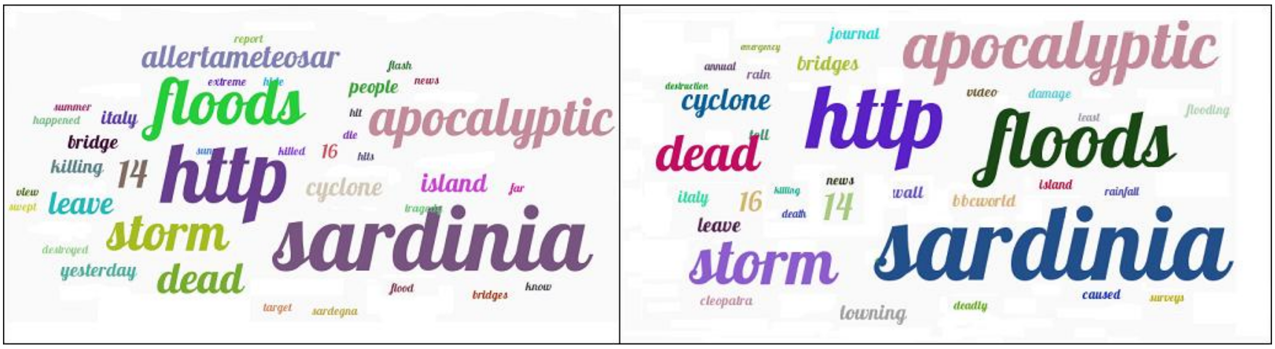


Figure 9. SMGI Textual analysis: Tag-clouds for Italy (left) and worldwide (right).

The results of tag cloud on worldwide tweets identify the words used for the data collection (Sardinia and floods), other words strictly related to the subject (storm, cyclone, cleopatra, rain, italy), several terms related to the consequences of the event (apocalyptic, dead, bridges), as well as, several potentially unrelated terms (wall, journal, bbc, http). The results of tag cloud on Italy tweets expose the same related terms, but also several different unrelated terms such as: allertameteosar and yesterday. A further investigation is conducted on the unrelated words worldwide, disclosing how they are caused by a viral sharing phenomenon of tweets posted by the Wall Street Journal and BBC World Twitter accounts. These tweets have caused the identified peaks of interest and the tag ‘http’ by means of their diffusion and their URL-related contents. At the same time, analysis results on Italian SMGI show the use of the word ‘allertameteosar’ as a tag used to identify the tweets related exclusively to the cyclone, while the word ‘yesterday’ is related to tweets that refer phenomena related to start of the hazard in Sardinia, such as afflicted area and accidents.

Finally, the investigation on Nigerian SMGI explains how the particular interest of the Nigerian social network community on a completely spatially unrelated topic is caused by the high sharing phenomenon, namely ‘retweet’, of news information among users, leaded by the Breaking News Nigeria Twitter account, which offers an overview of worldwide events. In conclusion, the provided results explain the causes of the peaks and demonstrate how the textual analysis may enhance the awareness about a topic from SMGI contents considering both the spatial, temporal and textual dimension of this particular source, strengthening the concept of media as GIS and the convergence of GIS and social media (Sui and Goodchild, 2011).

6.3 Example case study YouTube: the landscape perception

In the second example case study, the SMGI Analytics framework is applied on YouTube SMGI to analyze the perception of Sardinia local communities on a specific topic with relevance for urban and regional planning of Sardinia, namely 'landscape', investigating potential meaningful spatial patterns at the regional scale. The approach builds on a preliminary exploratory analysis of YouTube contents of Sardinia (Italy) related to the word 'landscape' in English language and 'paesaggio' in Italian language, investigating potential meaningful spatial patterns by means of a textual analysis on harvested contents, as well as, by integrating the SMGI dataset with A-GI from the Regional SDI. Indeed, the integration of A-GI with SMGI provided by users and local communities may disclose useful patterns and information both for design and decision-making processes.

6.3.1 Data collection

The data collection of contents from YouTube is performed by the YouTube Extractor tool, which allows the collection of SMGI from the social network by a natural language query bounded by a given radius around a specific location. For the purpose of this analysis the words landscape and its Italian translation 'paesaggio' are set as keywords for the extraction of videos in the Sardinian territory. The keywords for the query are set both in English and Italian to avoid the eventual dismissal of useful contents due to language constraints.

The resulting dataset contains 180 entries and allows the investigation of locations, activities and values, which may be considered tightly related to the 'landscape' topic according to user opinions. Moreover, the dataset includes several attributes obtained from the video metadata that concern the title, the author, the video description, the data of publication and the video URL, offering opportunities for the development of further analysis in combination with other spatial data layers.

6.3.2 Spatial-Textual analyses

The investigation of the contents is performed by means of tag cloud analysis on video title and description, classifying the words into 4 semantic categories, namely (1) toponyms/places, (2) activities, (3) values, (4) links/URL. In the first category the words referring to specific toponyms (city name) or physical places (urban location) are considered, while in the second category the words related to events or objects are grouped. The third category includes words related to adjectives that may be used by users for the personal evaluation of places and events, while the last category groups the words related to external links, blog or specific URL. The results of the tag-cloud and semantic analysis lead to identify the 10 most used words for each category, as shown in table 6 and figure 10.

SEMANTIC CATEGORY	WORDS [frequency]
Toponyms/Places	Sardegna [66], Terra [20], Hotel [18], Nuoro [13], Cagliari [13], Italy [11], Mare [10], Cabras [9], Parco [8], Cala [7]
Activities	Musica [10], Marcia [10], Foto [8], Servizio [7], Idee [7], Progetto [7], Pittura [6], Ecocampus [6], Festival [5], Valorizzazione [5]
Values	Dedicata [6], Realizzato [6], Bella [5], Patrimonio [5], Grande [5], Panoramica [5], Interno [4], Animato [4], Verde [4], Eolico [4]
Links/URL	http [87], youtube [70], www [44], google [24], mp4 [16], on [9], by [9], for [8], to [7], wmv [6]

Table 6. SMGI textual analyses: Top 10 words related to ‘landscape’ for each category.

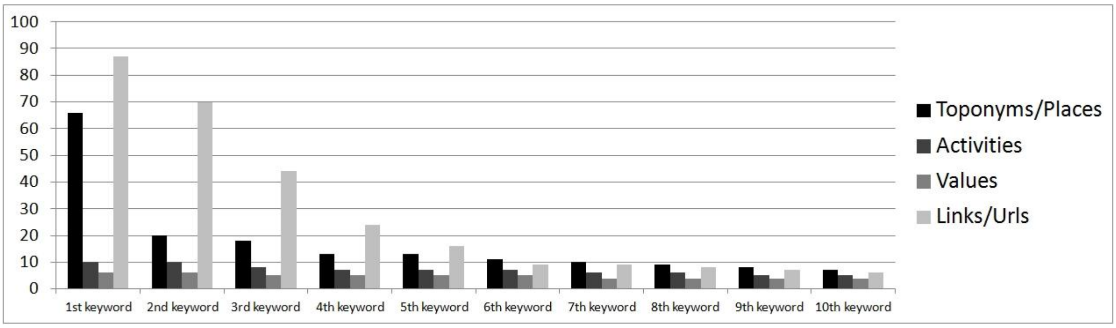


Figure 10. SMGI textual analyses: Top 10 words related to ‘landscape’ for each category.

Findings of tag cloud analysis and semantic analysis show how the words related to toponyms/places and links/URL exhibit high frequency, while the activities and the values category expose a constant trend.

Furthermore, spatial analyses are carried on to evaluate the spatial distribution of videos. The first analysis aims to investigate potential underlying spatial patterns or relationships between the spatial distribution of on landscape and the different land uses expressed by the CORINE land cover. The second analysis is performed to evaluate the percentage of videos occurring in the coastal area as defined by Regional Landscape Plan of Sardinia, firstly adopted in 2006. The goal of this analysis is the evaluation of where the thought ‘landscape’ of users is located in space.

The analysis of SMGI spatial distribution with regards to the CORINE land cover classes displays interesting results. The videos percentage for land class shows how 50% of videos fall in artificial surfaces areas (urban territory), but the explanation of this value may be easily found into two main reason. First, the YouTube service allows users to upload personal contents either in real-time or not, causing anomalies in geographic location that can be consequently shifted from the location of video to the location of user during upload. In this case, the location usually refers to the hardware platform in the urban territory.

Second, the development of several events related to the landscape in the urban territory, with massive participation of users may spread the value of uploaded videos in urban environments. Several questions could arise for the 15.56% of videos occurring in agricultural areas, but a large scale spatial analysis highlights both the inclusion of artificial grassland in this CORINE land use class, and the proximity of these emerging areas to the urban centers. At the same time, the percentage of videos occurring in forest and semi-natural areas, wetlands and water bodies may be explained considering the tight relationships between these areas and the term landscape.

The second spatial analysis considers the distribution of videos in the coastal area, exposing the following percentages: 58.33% for internal areas, 37.22% for coastal areas, 4.45% for sea areas. The outcomes display an almost uniform distribution at the regional level, but interesting results are obtained analyzing the spatial distribution for each Province and separately for each geographic area.

On the one hand, the Province of Cagliari provides 41.8% of videos for the coastal area and 6.67% for the internal area. Similarly the Province of Olbia-Tempio exhibits 19.4% for coastal area and 4.76% for internal area. On the other hand, the Province of Nuoro supplies 5.97% of videos for coastal area and 36.19% for internal area, while the Province of Oristano provides 22.38% for coastal area and 20.95% for internal area. These results may be considered significant of a different perception of landscape by users according to different geographic zone. Values for the Province of Cagliari and Olbia-Tempio may depict a high vocation toward the relationship of landscape with coastal area; in an opposite way the mountain Province of Nuoro displays an high vocation toward the internal areas landscape. The result for other Provinces presents a uniform pattern in the distribution of videos. The findings of this analysis are provided in figure 11.

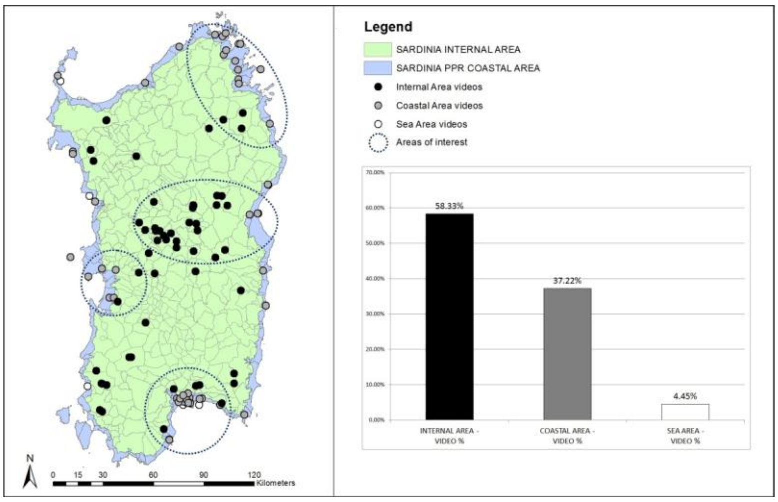


Figure 11. SMGI spatial analysis. Spatial distribution of video for Regional Landscape Plan areas.

6.4 Example case study YouTube: the neighborhoods perception

The third example case study deals with the investigation of the differences in perceptions as expressed by YouTube video publishers on several neighborhoods in the municipality of Cagliari (Sardinia, Italy) and for the 'Molentargius' Regional park. The aim of the analysis is the spatial and textual investigation of YouTube SMGI eliciting the differences in perceptions and opinions provided for the neighborhoods of *Castello*, *Marina*, *Is Mirrionis*, as well as for the *Molentargius* park area. These neighborhoods are chosen because of their physical, functional and social differences. *Castello* is the innermost historic central area, *Marina* is a commercial district that links the historic centre to the harbor of Cagliari, and *Is Mirrionis* is a community poor-housing area at the edge of the centre, built during the 1960s' city sprawl. The Regional Natural Park of *Molentargius*, instituted in 1999, is a wetland, neighboring heavily populated urban areas. The park offers an ideal habitat for several animal species including flamingo.

6.4.1 Data collection

The data collection of contents from YouTube is performed by the YouTube Extractor tool, selecting exclusively the geographic areas of interest in the municipality of Cagliari. The analysis purpose is the use of SMGI at the local scale in order to investigate the perceptions of users on neighborhoods and the park. The findings may have considerable importance as a support for the design, analysis and decision-making for interventions in the urban environment or in the protected area according to the contributed preferences. The data extraction results in a SMGI dataset of 385 observations, partitioned as follow: 101 video for *Marina*, 93 video for *Castello*, 84 video for *Is Mirrionis* and 107 video for the *Molentargius* Park. Similarly to the previous case study, the SMGI dataset includes the attributes concerning the title, the author, the video description, the data of publication and the video URL, offering opportunities for developing further analysis in combination with A-GI.

6.4.2 Spatial-Textual analyses

A tag-cloud analysis is conducted on the dataset in order to detect and classify the most significant words related to places, events and values for each neighborhood and for the park. The results offer a glimpse of the users' perception, showing information that usually is not provided in official land use documents. In this case a semantic analysis is conducted on the SMGI dataset exclusively to remove the words related to URL and external links, which are not able to provide useful information for depicting the users' preferences or opinions on the investigated areas. The results of the textual analyses, as well as, of the SMGI spatial distribution are displayed in Table 7 and Figure 12, respectively.

NEIGHBORHOODS	WORDS [frequency]
CASTELLO	Cagliari [48], Metropolitano [10], Sicuro [10], Concerto [6], Marmora [6], Palazzo [5], Regio [5], Storico [4], Cattedrale [2], Bastione [2]
IS MIRRIONIS	Cagliari [40], Torneo [10], Monteclaro [7], Cus [6], Calcio [4], FC [4], Sound [4], Finale [4], Music [3], Parco [2],
MARINA	Cagliari [61], Santa Lucia [13], Concerto [9], Capodanno [6], Festa [6], Sepolcro [5], Musica [5], Festival [4], Porto [4], Chiesa [4]
MOLENTARGIUS NATURAL PARK	Cagliari [62], Quartu Sant'Elena [15], Saline [10], Fenicotteri [8], Parco [8], Poetto [5], Conferenza [5], Servizio [5], Stagno [4], Monte Urpinu [3]

Table 7. Top 10 most significant words related to toponyms, events and values.

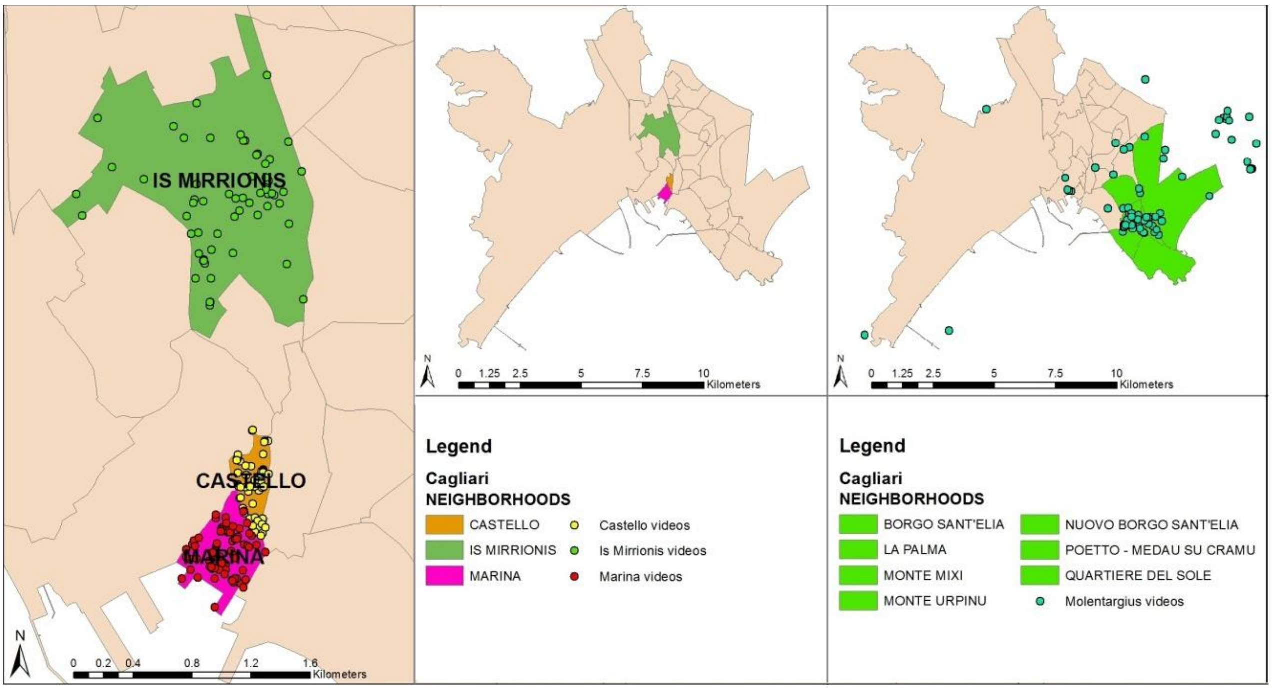


Figure 12. Spatial distribution of SMGI dataset for the investigated neighborhoods in Cagliari (Italy).

The results define a clear image of the overall perception of users for each considered area. Several keywords obtained by tag cloud analysis of title and description of videos depict the main features of the neighborhoods. The words related to *Castello* concern historical buildings, places, ways and typical events, while for *Marina* the words are related both to celebrations and specific buildings, such as church and harbor. Results of analysis on *Is Mirrionis* show a current habit of users to identify the district as supplier of football-related activities, and highlight the presence of the urban park of *Monteclaro*. In addition, the analysis on *Molentargius* area concerns the presence of the park, the several features of the place, and the specific toponym of *Quartu Sant'Elena*, that is another municipality surrounding the park.

The resulting overall opinion concerning the neighborhoods is particularly interesting due to its similarity with the results of another specific study developed on the same areas. Despite the differences in datasets, SMGI origin, users, and time periods, the findings obtained using the platform Place, I Care! (PIC) and Cagliari, I Care! (CIC) (Campagna et al., 2013), show similar results regarding users' opinions and interests, related to the examined neighborhoods. This phenomenon raises interesting questions on SMGI and on potential opportunities for analysis, and will be further investigated in future studies.

6.5 Example case study Instagram: investigating the geography of places

The last example case study proposes the SMGI Analytics application for investigating Instagram contents related to 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) shows 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) takes advantage of 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 application of SMGI Analytics on the case study of Iglesias municipality is carried out according mainly to the following three stages: (1) data collection, (2) spatial and temporal analysis, and (3) cluster analyses on SMGI dataset.

6.5.1 Data collection

The data collection of SMGI from Instagram is 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 results 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 to 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 are anonymized for privacy issues before the storage and processing. An exploratory analysis of the SMGI dataset exposes 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. The exploratory analysis results are shown in Figure 13.

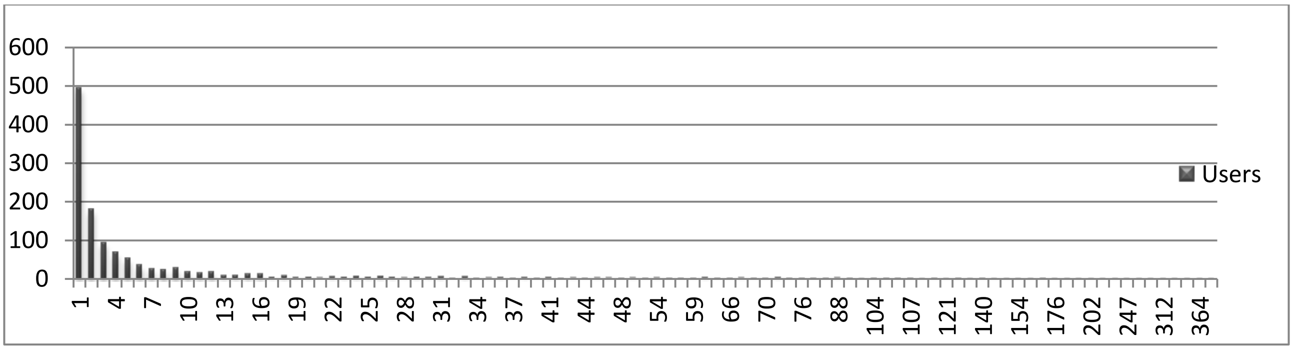


Figure 13. SMGI exploratory analysis. Users contributions: x-axis num. photo & y-axis num. users.

The findings show a notable heterogeneity among users contributions and expose an occasional access to the social network by one third of the users. Despite a different degree of participation by users, the dataset is investigated in order to identify similarities and differences among the contributions, as well as to investigate areas attracting high interest and urban dynamics.

6.5.2 Spatial-temporal analyses

After the data collection, the spatial and temporal components of the SMGI dataset are investigated directly in a GIS environment, in order to explore potential patterns of interest in the area and local community dynamics. At this stage, the SMGI dataset is integrated with several official datasets from the regional SDI of Sardinia related to the Iglesias municipality such as settlements, roads network, and buildings.

A simple investigation of the dataset spatial distribution shows 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 14.

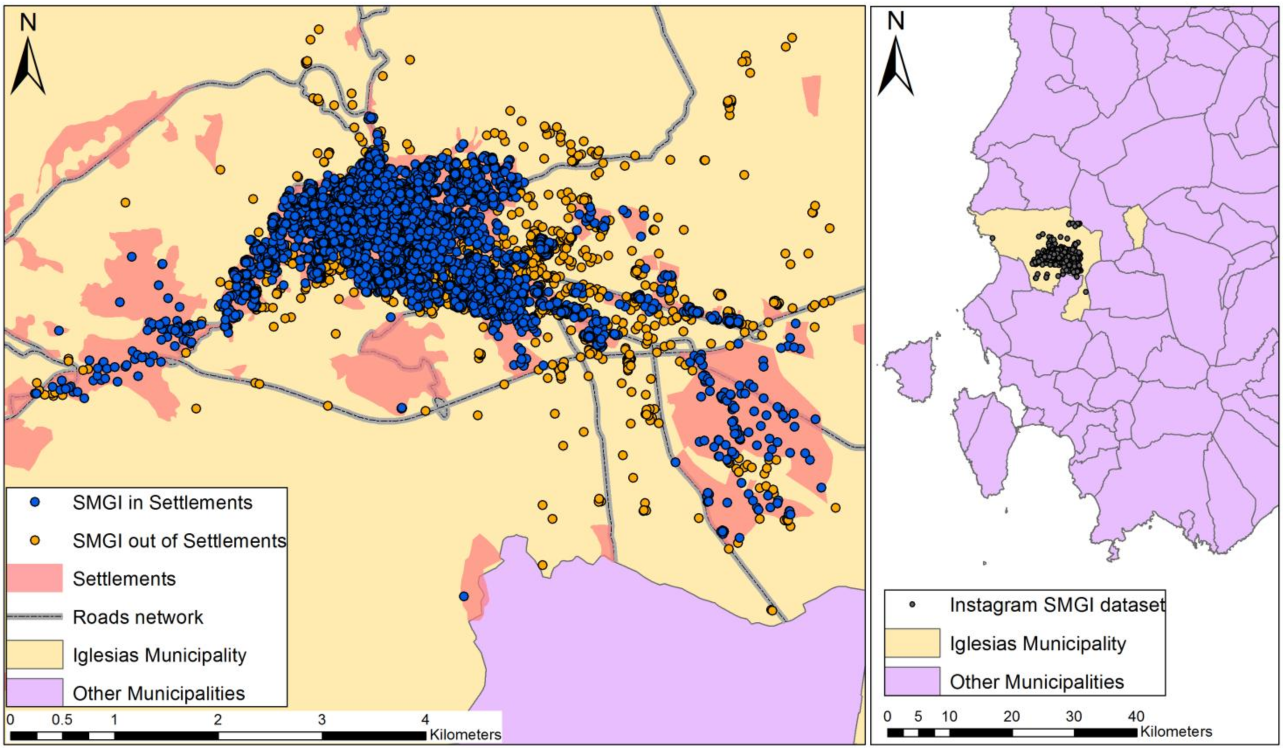


Figure 14. SMGI spatial analysis. Spatial distribution of Instagram SMGI in the Iglesias municipality.

The temporal component of the SMGI dataset is 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 is conducted investigating seasons, months, days of the week and hours of the day, disclosing interesting patterns. The results of temporal analysis show how SMGI is 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 is also evident in month distribution: July presents the highest percentage of produced contents (13%) and November the lowest one (5%).

The analysis of daily distribution provides more balanced results, with a slightly higher percentage of contents produced during weekends (Saturday and Sunday). Finally, the analysis of daily hours' trend allows identifying two main peaks of interest for both workdays (Monday to Friday) and weekends (Saturday to Sunday). The peaks are 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 stressing the use of the social platform during the meals or in break times. In contrast, the period 05:00-06:00 shows the lowest percentage of produced contents both for the workdays and the weekends. In spite of similar temporal peaks, the workdays and weekends trends expose 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 demonstrates how workdays and weekends' temporal patterns are similar for cities of the same country, but show major differences among cities in different countries. The results of temporal analysis for the different periods are provided in Figure 15.

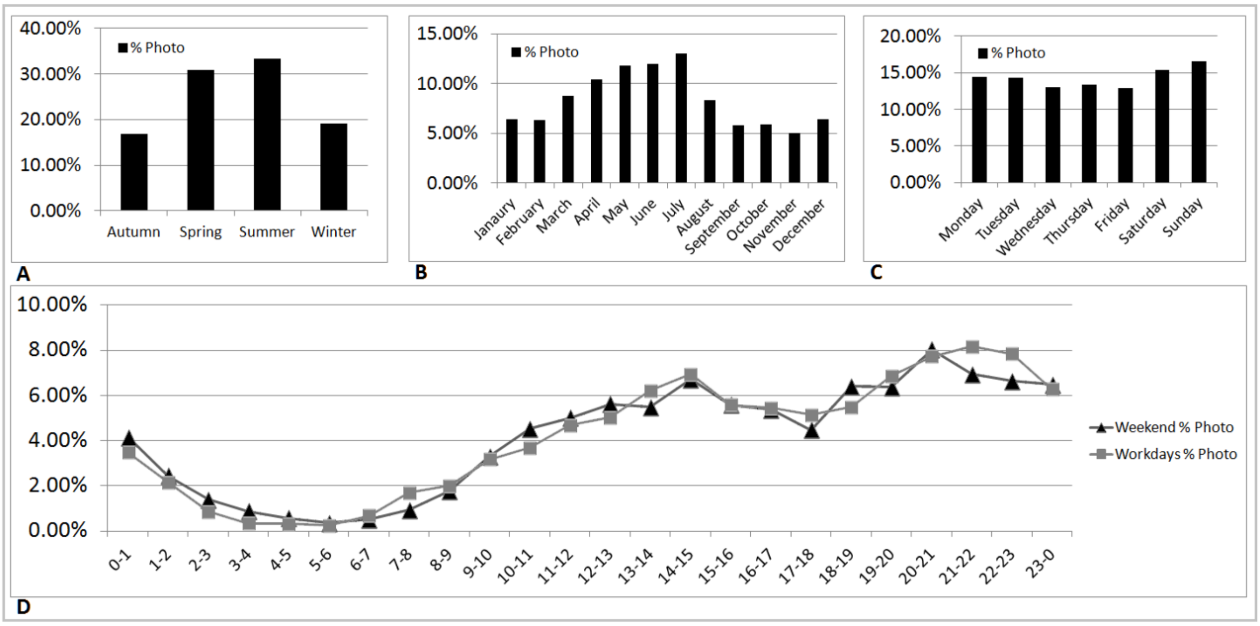


Figure 15. SMGI temporal analysis. Temporal distribution of Instagram SMGI dataset: (A) season, (B) month, (C) day of the week, (D) hours trend during weekends and workdays.

6.5.3 Cluster analyses

The results of spatial and temporal investigations lead 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 may foster the use of advanced analytical methods to identify, classify and interpret the users' interest toward certain specific spaces. For this purpose, the DB-SCAN algorithm and it slightly modified version called Feature-based DB-SCAN (FB-DBSCAN), integrated in SPATEXT, is used to compute clusters based on the spatial density of points.

As debated in the previous chapter, DB-SCAN 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 ε (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 is the identification of the places that attract the major interest of the local community, which in the research is proposed to be measured in terms of high density of contributions. Nonetheless, operatively there is no opportunity to establish the preferable value of ε and min_pts before the computation, therefore the DB-SCAN tool is 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 lead toward the identification of the following values, which have proved to be preferable for the purpose of the study: ε = 20 meters and min_pts = 5. Indeed, the ε value, or threshold distance, is able to cover the dimension of a medium-sized fabric, while the min_pts value is 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 enable 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 emerge from the analysis, identifying the areas attracting the highest interest by users within the urban context. These areas may represent public spaces, private places, residential and commercial areas. The cluster analysis results are shown in Figure 16.

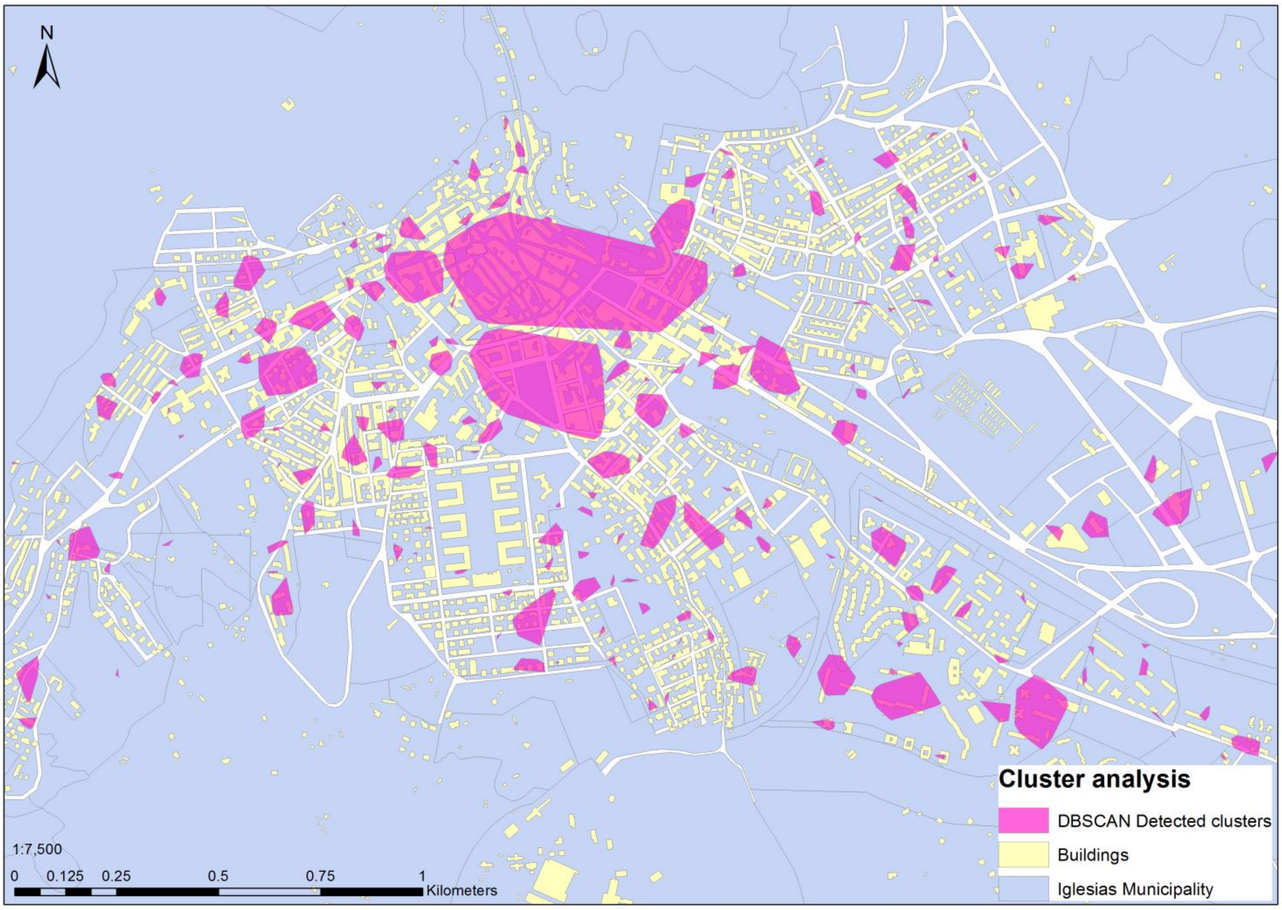


Figure 16. SMGI cluster analysis. Clusters identified by DB-SCAN.

Nonetheless, in order to find an explanation at the two clusters deserving the major attention of users in the urban environment of Iglesias, a further SMGI extraction is conducted by collecting data from the social networks Foursquare and Instagram Places, respectively. Both these social platforms are location-based social networks, which may provide useful information about POIs in the municipality.

The extraction results are compared in order to detect the common POIs in both the platforms and then to identify the 5 most visited places in the major clusters by means of an overlay analysis. The integration of complementary SMGI dataset demonstrate how these areas concern both the historic center of Iglesias and public space areas. A closer look to the clusters shows that the top cluster includes the historic Cathedral of *Santa Chiara*, the main avenue for leisure and night life of the municipality, namely *Via Matteotti*, and two of the main squares of Iglesias. At the same time, the bottom cluster contains the public park of the municipality. The results of the SMGI complementary integration are shown in next figures. The Figure 17 identifies the 5 most important POIs detected in the clusters.

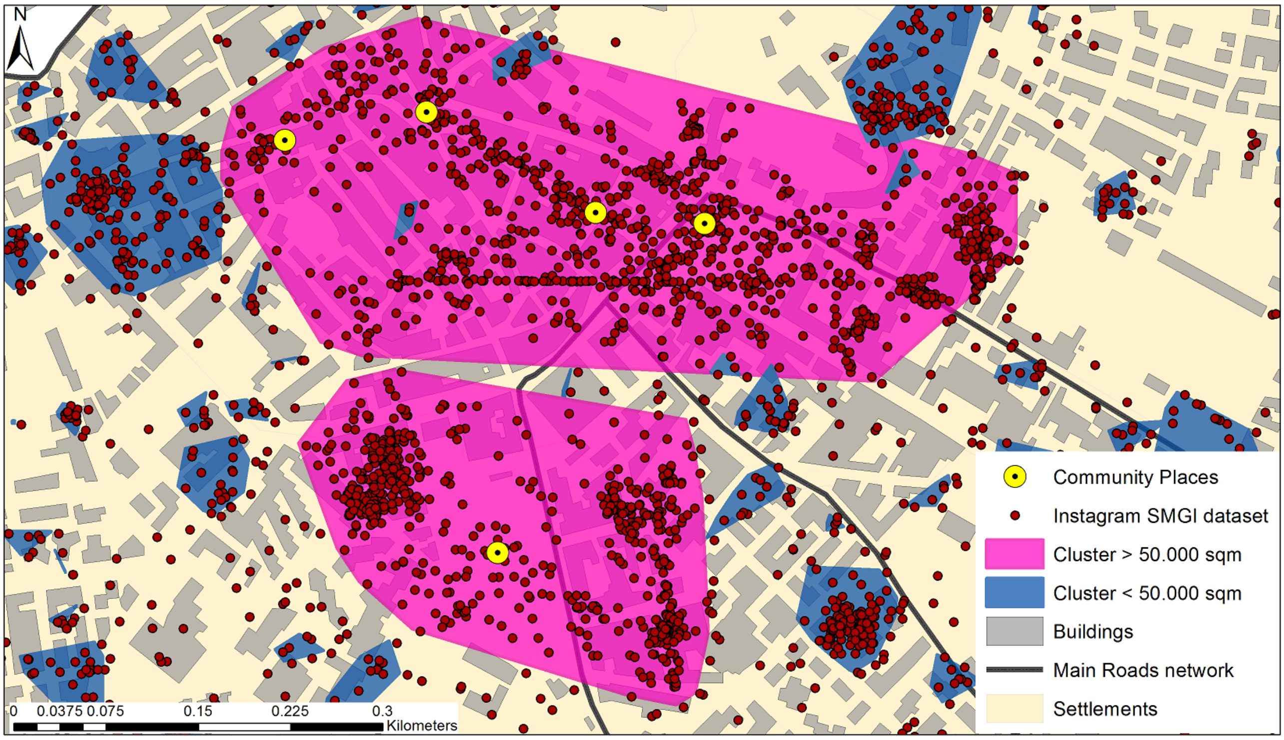


Figure 17. Identification of POIs in the major clusters.

In addition, the next figures 18, 19, 20, 21 and 22 show the identification of the Cathedral of Santa Chiara, the square Piazza La Marmora, the avenue Via Matteoti, the square Piazza Sella and the municipal public garden, respectively. The proposed approach demonstrates how the use of complementary SMGI may be useful to elicit information on places in near real-time, reducing the issues for direct investigations or surveys conducted on the local community. Obviously, the findings of the approach might be further investigated with traditional methods; however, this kind of exploratory analysis may be useful for informing and guiding further analytical efforts.

Figure 18. Identification of the Santa Chiara Cathedral.

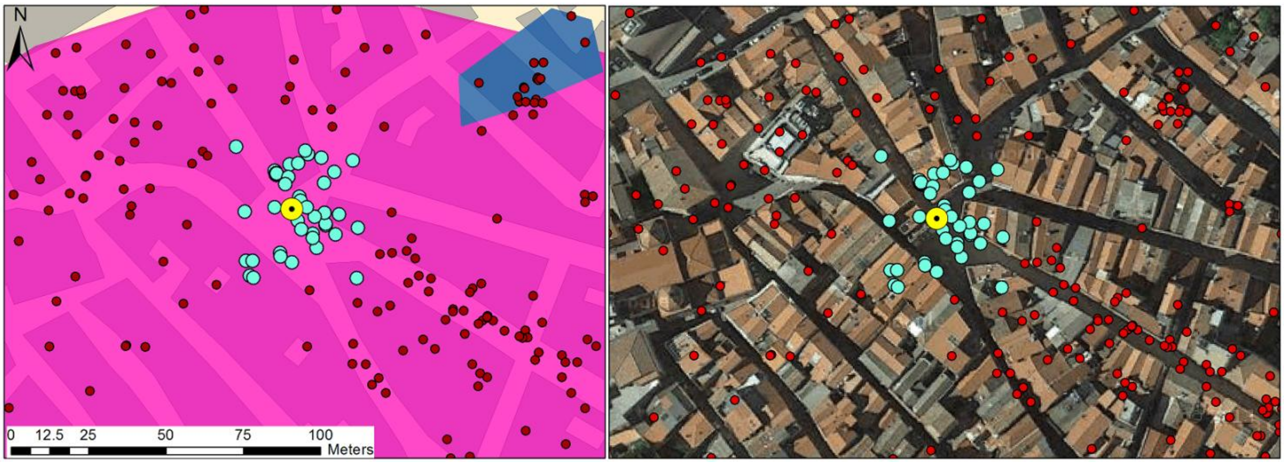


Figure 19. Identification of the Piazza La Marmora square.

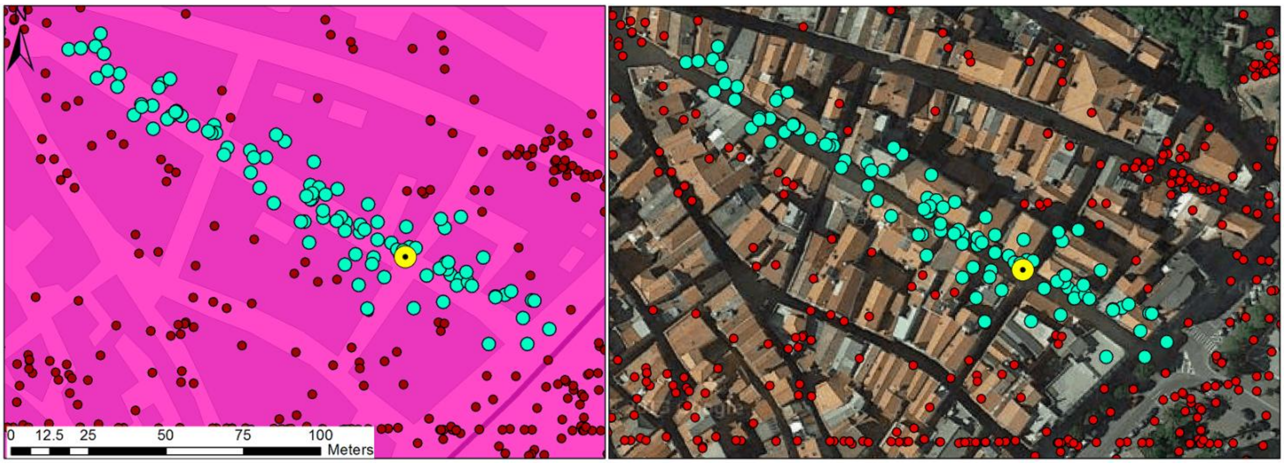


Figure 20. Identification of the Via Matteotti avenue.

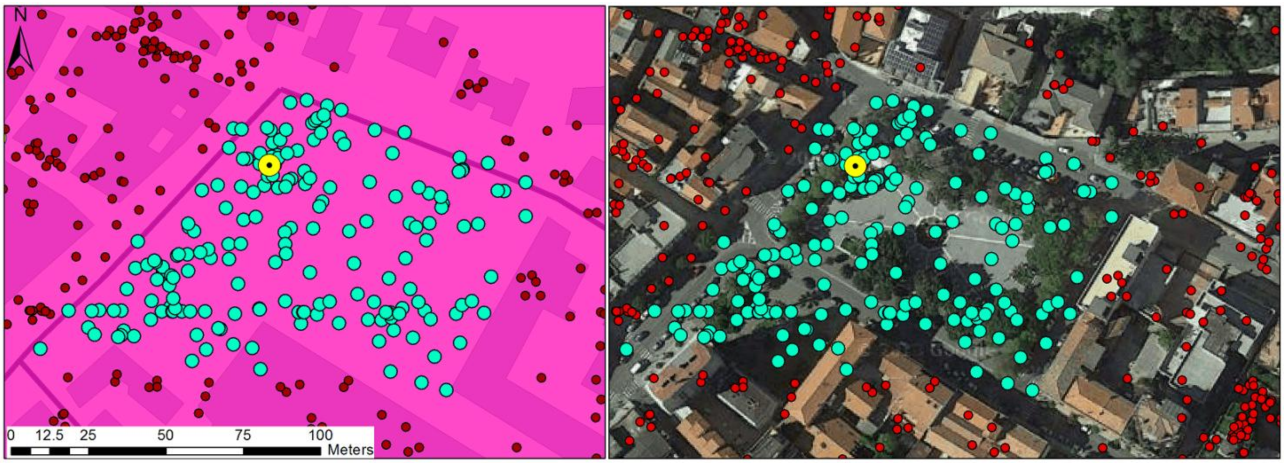


Figure 21. Identification of the Piazza Sella square.

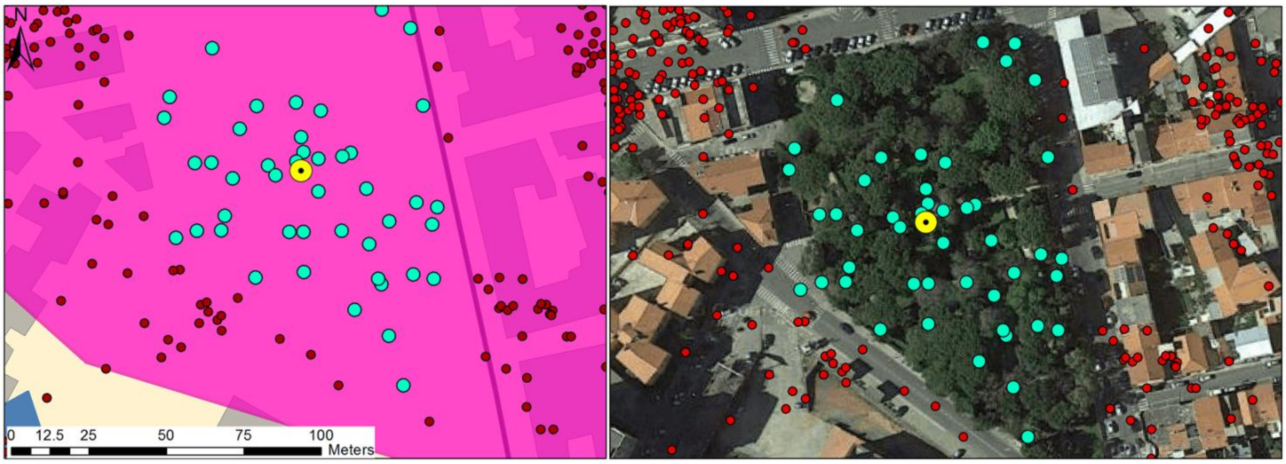


Figure 22. Identification of the municipality public garden.

Along the same vein, the FB-DBSCAN tool is 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 identify 368 clusters concerning 266 users. In this case, the number of identified clusters is higher than the value obtained in the previous analysis by means of DB-SCAN, but the clusters' sizes is notably smaller, focusing on specific places or fabrics within the municipality. 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 results of the clustering analysis performed through the SPATEXT tool FB-DBSCAN are shown in Figure 23.

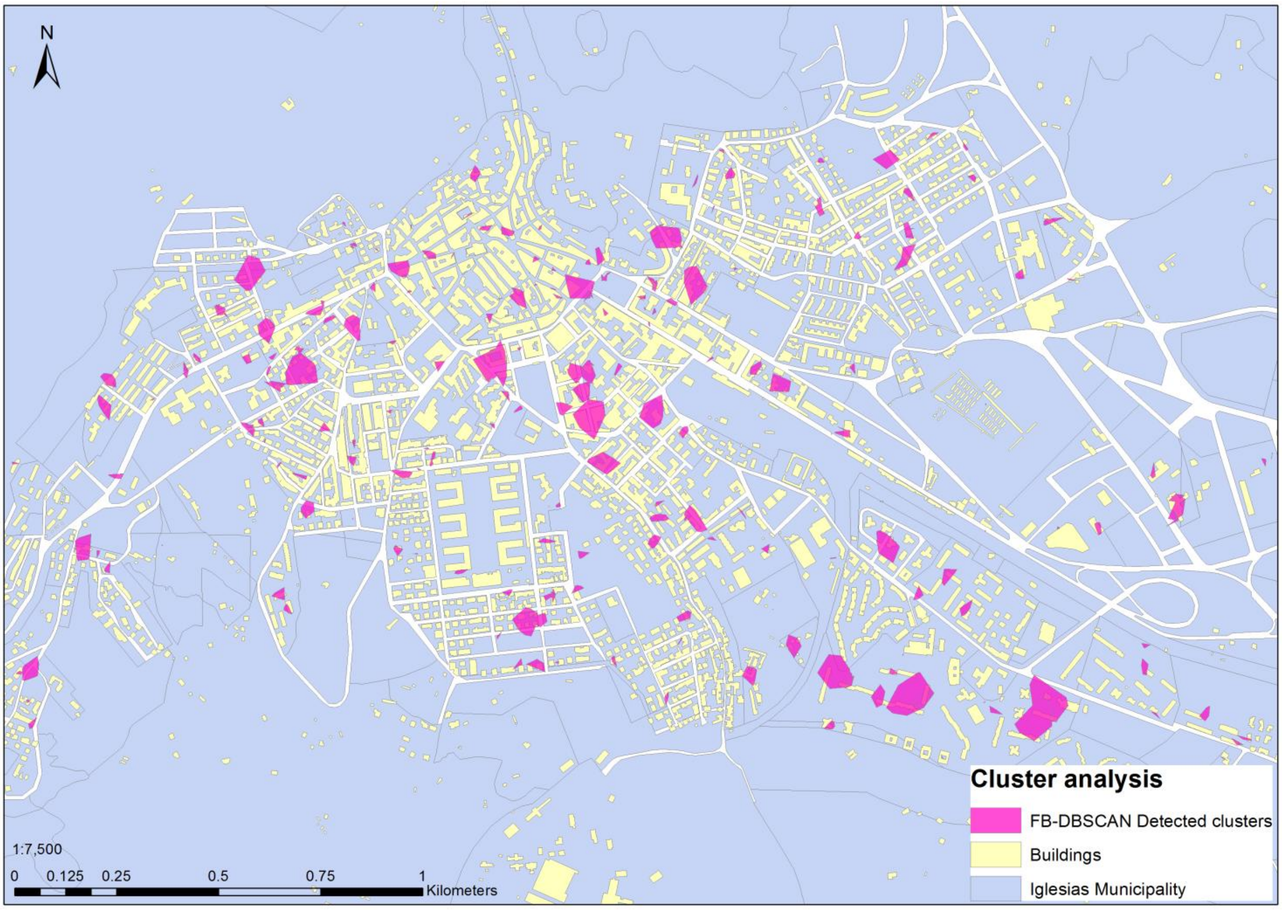


Figure 23. SMGI cluster analysis. Clusters identified by FB-DBSCAN.

The real clusters utilization by users may be discovered analyzing several parameters related to spatial and temporal characteristics, as well as by integrating further spatial information. In this case, the aim of the study is twofold: 1) identify the users' residential location, and consecutively 2) detect eventual not mapped buildings in the official information. In order to accomplish the analysis, the latest official buildings dataset from the Regional SDI is integrated.

This official dataset is selected in order to check the consistency of the clusters' location with the urban fabrics, easing 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.

The approach proposed in the research identifies six different parameters concerning the cluster itself or the included contributions, as described in Table 8.

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 8. Parameters used to identify residential clusters.

The above parameters are strictly related to the nature of the Instagram SMGI dataset and are selected specifically in order to ease the identification of dwelling areas from the single user clusters. Nevertheless, the parameters may be modified in order to search for other types of objects within any set of clusters. As a matter of fact, the Instagram SMGI data model provides information related to different dimensions, namely spatial, temporal, user and textual, paving the way for the identification of other sets of parameters, suitable to identify several urban uses and locations, such as night locals, schools, commercial services or working places, to name a few. For example, in order to identify night locals, the set of parameters might be modified to include exclusively clusters containing SMGI posted mostly during the night hours and by an established minimum number of users. Similarly, the working locals might be identified selecting clusters that show most of the contributions within specific time intervals during workdays and a lack of contribution during offices closure days. The temporal patterns of SMGI may be considered representative of different urban uses and in literature many studies rely upon the temporal dimension of contributions to investigate the urban land uses (Frias-Martinez et al., 2012; Torres and Costa, 2014; Silva et al., 2013 B).

The values of the six parameters are estimated for each cluster, while several combinations of the values are iteratively evaluated in order to identify exclusively the residential clusters. The following set of values results as the most appropriate to classify a cluster as residential in the study area: Cluster Centroid and Contributions Centroid have to be 1 (yes), while Number of Contributions and Time Span Among Photos have to present the highest values among clusters of the same user, or the values have to be higher than 10 and 30, respectively. Finally, Cluster Intersections has to be equal or lower than 2, while Cluster Density has to be higher than 4. The above parameters allow the identification of 47 residential clusters, which are confirmed by an overlay analysis with satellite imagery in GIS environment.

Afterwards, the same set of parameters is 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, may be considered as a sort of residential fabrics footprint among clusters and are used for the investigation. The parameters' values used to identify the missing buildings in A-GI are listed in Table 9.

Parameters	***Description***	***Value***
Cluster Centroid	The overlap of the cluster's centroid with an official building footprint is estimated	0 [boolean]
Contributions Centroid	The overlap of the cluster's contributions centroid with an official building footprint is estimated	0 [boolean]
Number of Contributions	The total number of contributions contained in the cluster is estimated	> 10 contributions
Cluster Intersections	The total number of intersections between the cluster's shape with other clusters	<= 2
Cluster Density	The ratio between the cluster's area and the number of contained contributions	=> 4 sq. meters
Time Span Among Photos	The time passed between the first contribution and the last one in the cluster	> 30 days

Table 9. Values of parameters for identifying residential clusters.

The analysis identifies 40 clusters, which are then visually assessed through satellite imagery to confirm the presence of not mapped buildings in A-GI. The visual assessment detects 9 not mapped buildings. Nevertheless, at the same time the other 31 identified clusters are confirmed as residential areas, but in this case the buildings are already mapped in A-GI. This issue may depend upon the lack of tolerance in the analysis performed during the comparison of Cluster Centroid and Contributions Centroid parameters with the official buildings dataset. An example of the analysis results is provided in Figure 24, wherein 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.

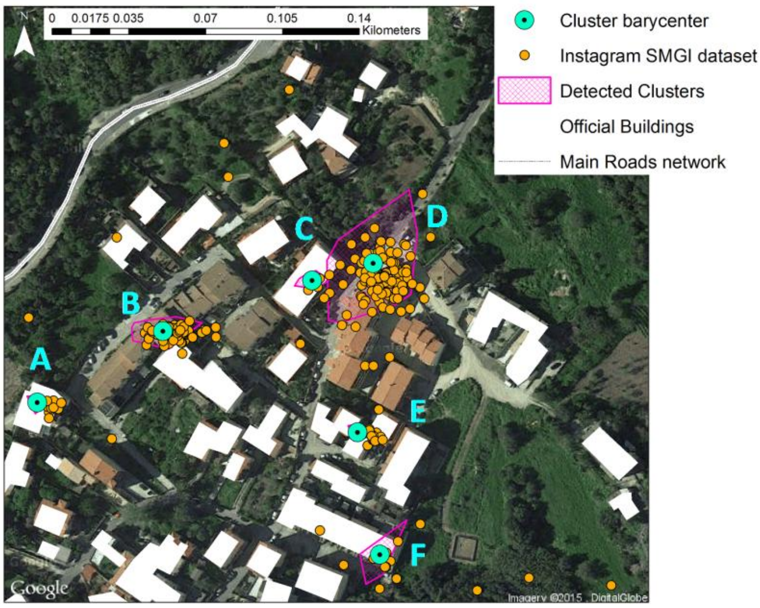


Figure 24. Results of clusters investigation in the Iglesias municipality.

In this example, the manual investigation through the Google Maps satellite image enables the detection of two buildings which are not mapped in the 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.

6.6 Discussion

This chapter discusses a number of example case studies carried on at different geographic scales, in order to investigate both the local communities' perceptions on relevant topics for spatial planning and the geography of places. Altogether, these examples contribute to demonstrate how SMGI may be used to elicit information, not only about the physical geography of places, to integrate existing A-GI, but also to express the perceptions of places and issues in time and spaces by the involved community, adding a pluralist perspective for spatial planning and decision-making processes. As a matter of fact, SMGI may be integrated with A-GI and used to understand people perceptions, contributing to define a pluralist model of local identity. The results of these first experiences offer an overview on potential uses of SMGI to investigate what people observe, evaluate, and how they behave in space and time. However, the underlying endeavor of the case studies is to test both the novel SMGI Analytics framework and the SPATEXT tools, which may ease the access to this novel data source to extract meaningful knowledge for spatial planning. The discussed case studies confirm the proposed assumptions concerning the SMGI Analytics framework and the SPATEXT tool. The novel methodological approach and the ad-hoc developed instruments may be able to foster the use of SMGI into practices, easing the integration of this experiential knowledge with the available official information.

The novel clustering approaches for identifying areas and POIs are evaluated in practices, demonstrating to be appropriate to deal with SMGI, enabling the elicitation of useful information. What appeared clear from the development, as well as, from the usage of SPATEXT, is that any SMGI source is characterized by a specific data model and by different rates in geographic diffusion. Therefore, different combinations of analytical approaches are required to interpret the results and consecutively the local context, appropriately. However, exploiting SMGI from different popular social media platforms demonstrate the capability of these sources to provide interesting results both for detecting changes in topography, as in the case of Instagram photos, as well as for investigating users perceptions and opinions, as in YouTube and Twitter cases. Finally, the discussed example case studies are useful to formalize the different stages of the SMGI Analytics framework, which is then applied on a complex case study strictly related to spatial planning analysis and decision-making in the municipality of Cagliari in Sardinia.

CHAPTER 7

SMGI ANALYTICS FRAMEWORK: PUBLIC SPACES ANALYSIS

7.1 Introduction

The results achieved from the early case studies demonstrate that the wealth of information enclosed in SMGI might be proficiently used to investigate the concerns and the attentions of people toward places, as well as, the users' behaviors and movements in space and time, fostering opportunities for gaining insights about urban dynamics and users preferences. Nonetheless, the case studies stress how different analytical approaches, integrating multiple social networks data with official information, may be required in order to investigate the local contexts appropriately. Several SMGI Analytics stages were tested and evaluated dealing with SMGI from different sources and at different geographic scales, exposing a number of potential approaches to elicit knowledge in order to explain examined phenomena. However, the obtained findings highlight the differences arising from the use of different social media platforms for developing specific spatial analyses.

The Instagram social network allows the extraction of massive georeferenced data for any time period and location and exposes a wide diffusion among users in the Sardinia Region (Italy). Therefore, this social media platform appears as the most suitable to enable SMGI Analytics in order depict the users' dynamics and preferences, enabling the investigation of spatial phenomena in urban environment at the local scale. Moreover, the point data model of Instagram SMGI is suitable for developing novel clustering methods, which may take into account the spatial, temporal and user distribution of contributions, easing the investigation of the places' geography. As a matter of fact, the proposed clustering analyses, relying on the SMGI density, may allow the identification of public and private areas of interest, as well as, of specific POIs in the urban environment.

In the light of these considerations, this chapter discusses the SMGI Analytics framework application on an Instagram SMGI dataset in order to investigate the public spaces of the Poetto beach and the Regional Park of Molentargius in the municipality of Cagliari, Sardinia (Italy). Operatively, the study is based upon all the stages identified for the SMGI Analytics framework, although the opportunities arising from textual analyses are strongly limited by the lack of structured textual contents in Instagram SMGI, and the step is neglected in this study. Actually, the case study is carried out through the following steps:

1. data collection;
2. explorative spatial-temporal analyses and A-GI integration;
3. spatial-temporal-user cluster analyses;

4. geodemographic classification;
5. user profiling;
6. multi-dimensional analyses on public spaces.

In particular, the analysis of SMGI user component plays a more important role in this study, providing opportunities to analyze the urban dynamics and preferences of specific users groups in space and time. Hence, an ad-hoc geodemographic classification (Webber and Craig, 1978; Sleight, 1997) of the Sardinian territory is conducted, relying on the official census data provided by the ISTAT. The geodemographic classification findings are then discussed and integrated with the clustering analyses results to achieve the Sardinia population profiling. Thus, the final results are used for developing a number of multi-dimensional analyses regarding concerns and preferences of the different identified population groups in several locations within the study area.

7.2 Data collection

The data collection is carried out by the SPATEXT Instagram extractor tool setting the spatial query on the public spaces of the Poetto beach and the Regional Park of Molentargius, and the temporal query on a one year period (26 January 2014 to 25 January 2015). The data extraction results in a one year sample of 34,776 geotagged photos from 8,350 users concerning areas in the Cagliari municipality and partially in the surrounding the Quartu Sant'Elena municipality. The Instagram SMGI point feature dataset is automatically created and each photo is georeferenced accordingly to the spatial information embedded in the geotag provided by the user during the upload on the Instagram service.

In addition, the SMGI dataset is enriched with a set of attributes made freely available by the Instagram API, such as: name of the place where the photo has been taken, user name, user ID, user picture URL, multimedia content URL, multimedia content's date of creation, number of comments, number of likes, multimedia content's caption and tags. Notwithstanding the public availability of personal users' data, the SMGI dataset is anonymized before the storage and the processing in order to ensure the respect of users' privacy.

An exploratory analysis of the SMGI dataset shows heterogeneous rates of contributions by the users. Indeed, 43.85% of users contributed with only 1 photo during the extraction period, while 48.52% of users shared between 2 and 10 photos. The 7.37% of users' uploaded photo to the social network with contributions ranging from 11 to 80, meanwhile exclusively the 0.26% of users provided more than 80 photos in the one year period. The contributions rates are shown in Figure 25.

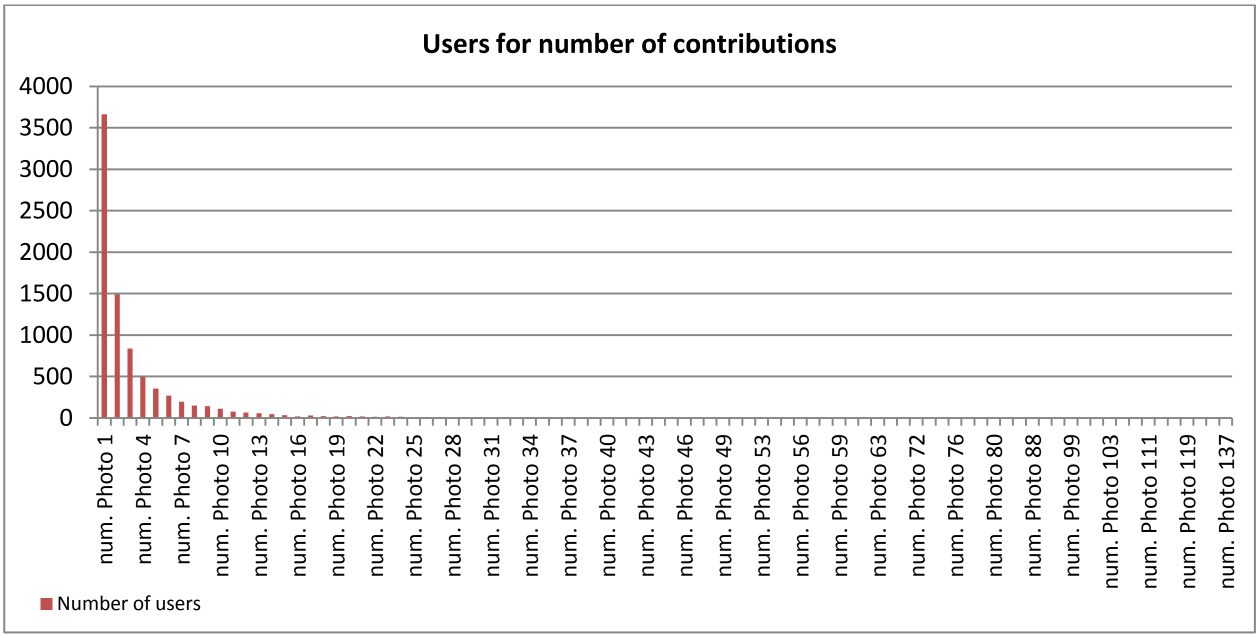


Figure 25. Users contributions: x-axis number of photo - y-axis number of users.

The strong heterogeneity exposed by users' contributions rates, and particularly the users with a limited number of contributions, may cause issues for the public spaces analysis, as well as, restrict the opportunities to develop a proper user profiling. Hence, the study proposes a second SMGI extraction from Instagram in order to deal with this issue. The second data collection concerns the extraction of all the contributions provided by the same users for the same time period, but without geographic boundaries.

This way, all the photos uploaded to the Instagram service by the users are extracted, potentially overcoming the hurdles related to users with insufficient contributions, and easing both the investigation of users' dynamics and the development of cluster analyses to identify the places attracting the highest individual attention.

The second Instagram SMGI dataset contains over 840K photos produced by 8,159 users. In this case, the number of contributing users is smaller due to a change in the privacy settings of 191 users, which forbid the extraction of personal contents from the Instagram API.

An exploratory analysis of this dataset exposes the following contributions rates: the 1.39% of users contributed with only 1 photo, the 10.31% of users shared between 2 and 10 photos, the 49.30% of users provided contributions ranging from 11 to 80, while the 39% of users contributed with more than 80 photos to the dataset. The second SMGI dataset contributions rates are shown in Figure 26, considering ten ranges.

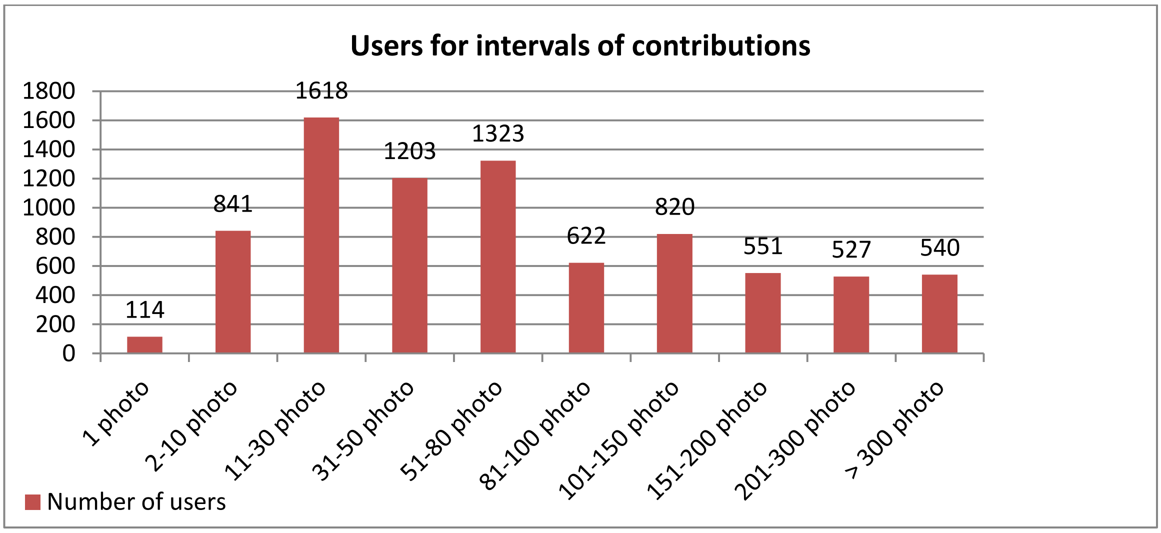


Figure 26. Aggregated users contributions: x-axis range of contributions - y-axis number of users.

The depicted values demonstrate how the second extraction strengthens the opportunities to develop cluster analyses in order to identify the potential users' residential location from the personal contributions spatial pattern. As a matter of fact, about 88% of users contributed with more than 10 photos, easing the data processing in search of interesting personal clusters at the global scale.

7.3 Explorative spatial-temporal analyses and A-GI integration

Following the data collection, the spatial and temporal dimensions of the Instagram SMGI datasets are investigated directly in GIS environment, searching for potential patterns of interest. The first SMGI dataset, concerning the public space areas of Cagliari and Quartu Sant'Elena municipalities, is investigated both in terms of spatial and temporal distribution. In addition, it is integrated with several A-GI from the Regional SDI, such as settlements, roads network, buildings, parks and protected areas. As a matter of fact, on the one hand the integration of voluntary and official information is a fundamental step to prepare the ground for further analyses and to eventually assess the SMGI quality and reliability (Spinsanti and Ostermann, 2013).

On the other hand, the second SMGI dataset investigation is limited to the spatial distribution analysis in order to visually explore the geographic extents of users' contributions. The spatial distributions of the two Instagram SMGI datasets are provided in figure 27 and figure 28, respectively.

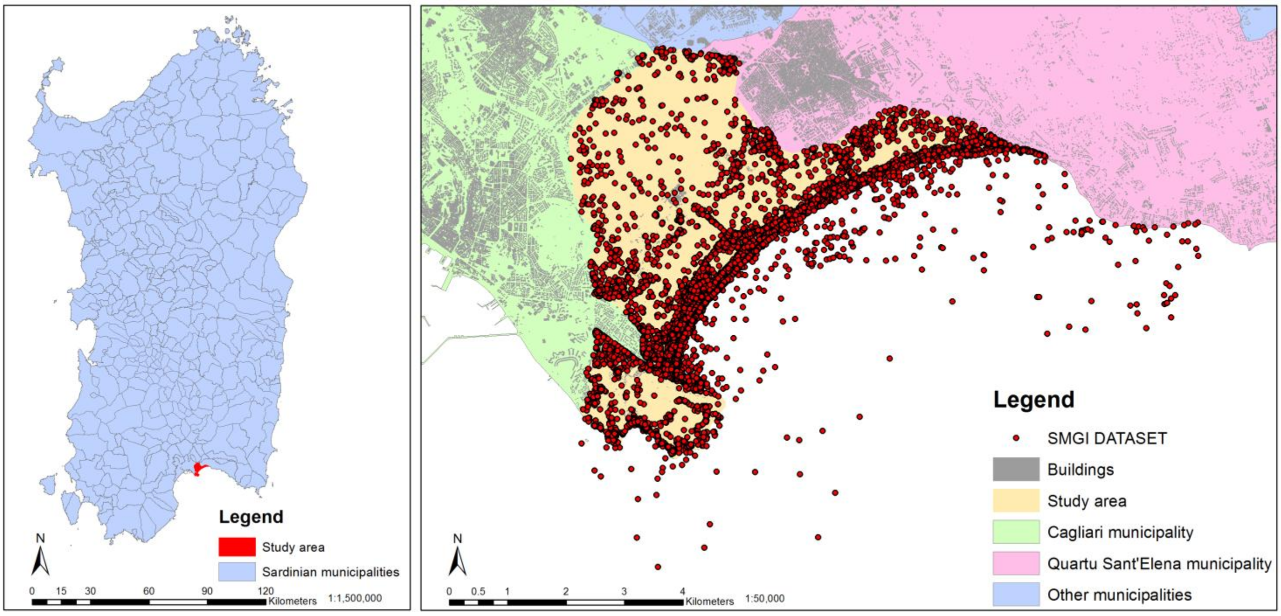


Figure 27. Spatial distribution of SMGI related to public spaces in Cagliari and Quartu Sant’Elena.

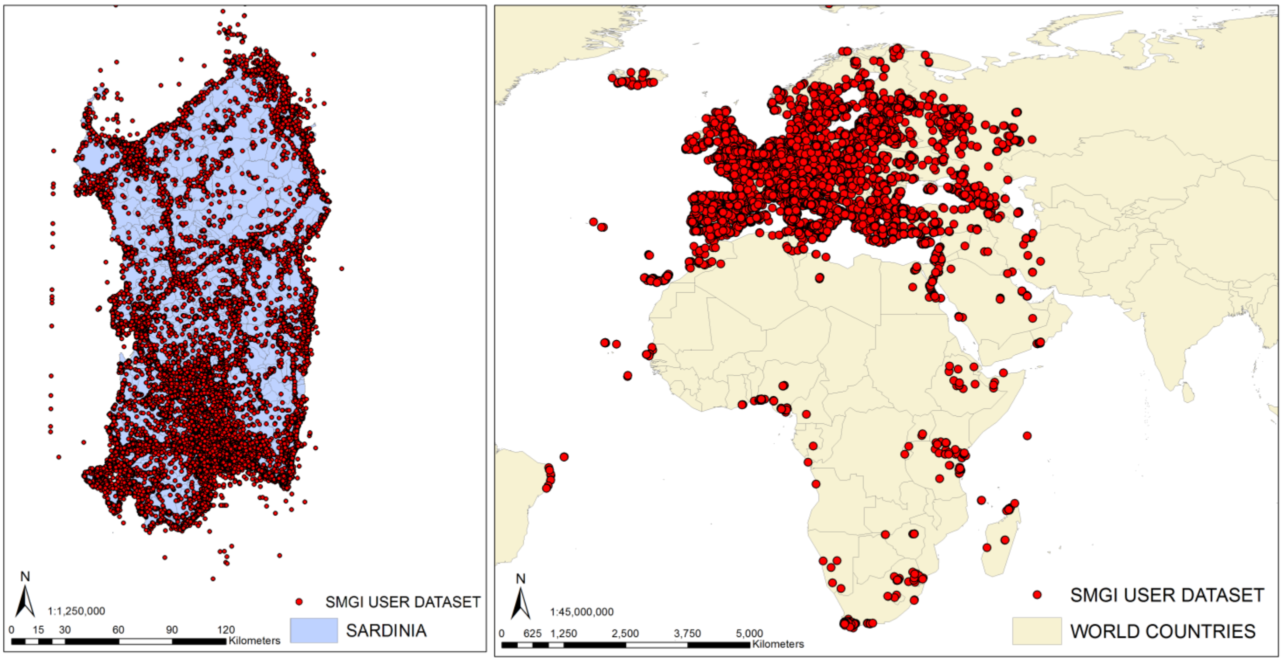


Figure 28. Spatial distribution of SMGI concerning the users’ contribution worldwide.

The proposed figures show the extents of users’ contributions both in the areas under examination in the case study and across the world countries. The second dataset’s notable extents demonstrate the wide diffusion of the Instagram social network and may favor the investigation of users’ movement worldwide, fostering the identification of personal places of interest.

Following the explorative spatial analysis, the temporal component of the first SMGI dataset, related to the areas in the Cagliari and the Quartu Sant’Elena municipalities, is investigated for different periods searching for peaks of interest, trends and patterns of interest. The explorative temporal analysis is carried out concerning the patterns of seasons, months, days of the week as well as the workdays and weekends’ trends. First of all, the results of temporal analysis related to seasonal patterns show how SMGI in examined areas is increasingly produced and uploaded during the summer (44.07%) in the period from June to August. Then, 22.16% of contributions are provided during autumn (from September to November) and 19.57% during spring (from March to May). Finally, during winter (from December to February) the SMGI dataset provides the lowest number of contributions (14.20%). The result of the seasonal-temporal analysis is provided in Figure 29, exposing the contributions rates for each period.

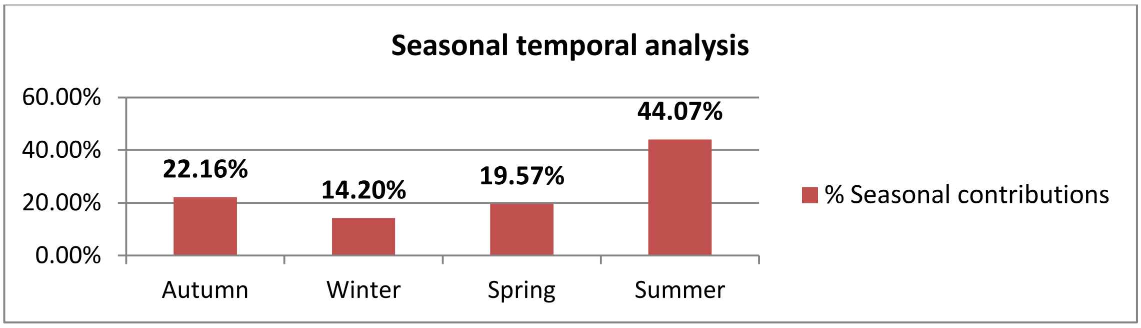


Figure 29. Seasonal SMGI percentage.

Secondly, the temporal analysis concerns the investigation of the SMGI dataset monthly patterns. The results show that August is the month exposing the major number of contributions (16.91%), followed by July (13.85%) and June (13.32%). Conversely, the months with the lowest volume of SMGI are December (4.30%), February (4.66%) and March (4.43%). The monthly temporal trend for the examined areas starts to rise from April (7.36%) until August (16.91%), then from September (10.41%) it steady decreases till December (4.30%). A slight increase is present in the SMGI volume during January (5.24%), but the trend immediately drops, as shown in next figure (Figure 30).

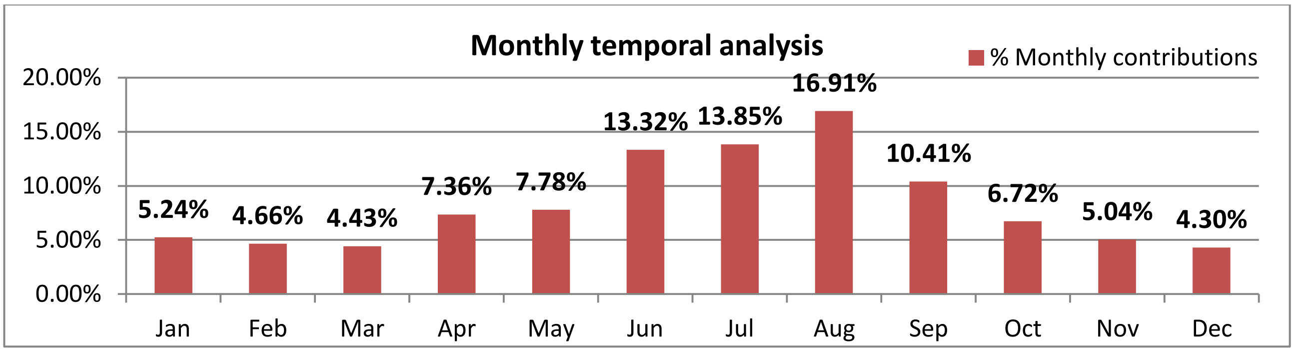


Figure 30. Monthly SMGI percentages.

The temporal analysis of the SMGI dataset's weekly patterns shows a steady trend from Monday (13.25%) to Friday (12.51), while a gradual increase starts from Saturday (17.70%) until Sunday, which exposes the highest number of contributions (21.55%). This temporal pattern demonstrates the major use of the

Instagram service by the users during the weekends probably due to leisure activities. The analysis findings are shown in Figure 31.

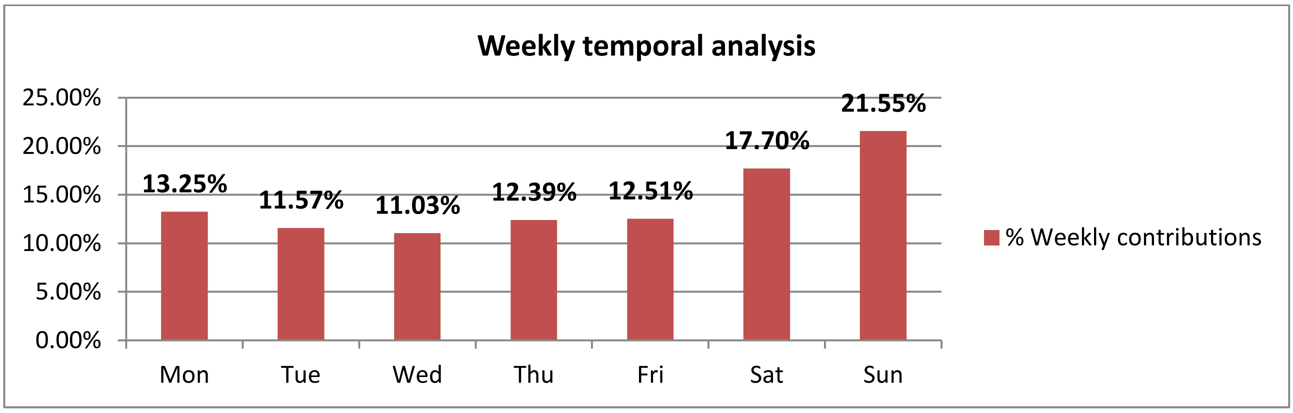


Figure 31. Weekly SMGI percentages.

Afterwards, the temporal analysis investigation concerns the daily trends of SMGI contributions with the aim of identifying users sharing behaviors. In addition, the same analysis is conducted also for workdays and weekends separately in order to detect potential differences in sharing patterns. The analysis of daily hours trend shows a gradual increase from the morning hours (06:00-07:00 period) when the shared volume is minimum (0.58%) to the 14:00-15:00 period, when the volume of shared SMGI is the highest (7.86%). Then the trend goes down, notwithstanding a plateau between the 16:00-17:00 PM period (6.54%) and the 20:00-21:00 period (6.38%).

Overall, the findings demonstrate that most of the contributions are shared from the afternoon until the evening, as shown in Figure 32. At the same time, the analysis is carried out both for workdays (Monday to Friday) and weekends (Saturday to Sunday) in order to identify potential discrepancies. Nevertheless, the results provided in Figure 33 dismiss the presence of notably differences between trends, confirming the highest volume of sharing in the 14:00-15:00 period meanwhile the lowest one is in the period 05:00-06:00 for workdays and 07:00-08:00 for weekends. The provided results might represent a descriptor of the typical cultural behaviors of the users or a sort of cultural footprint of the place that is mainly lived during specific temporal periods.

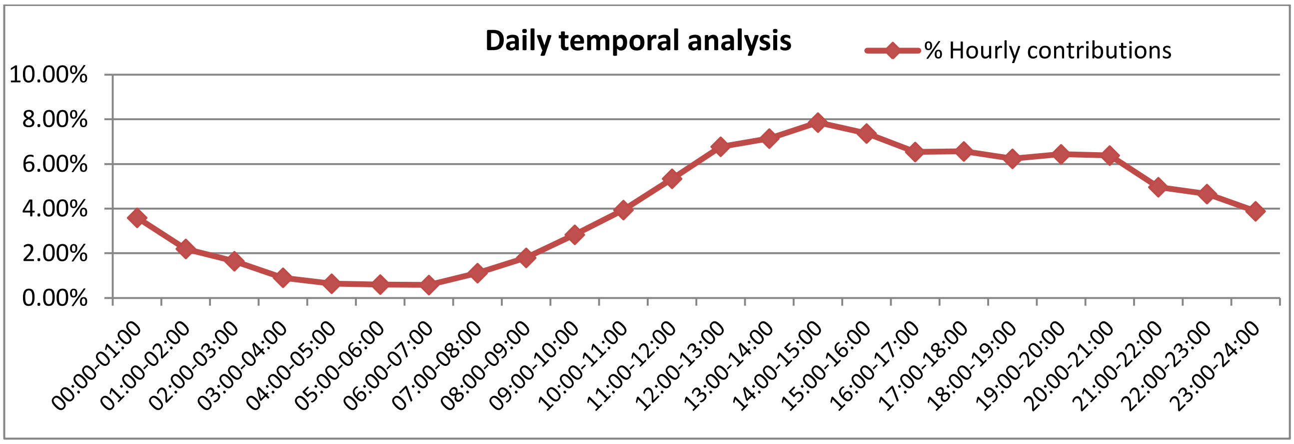


Figure 32. Daily SMGI percentages.

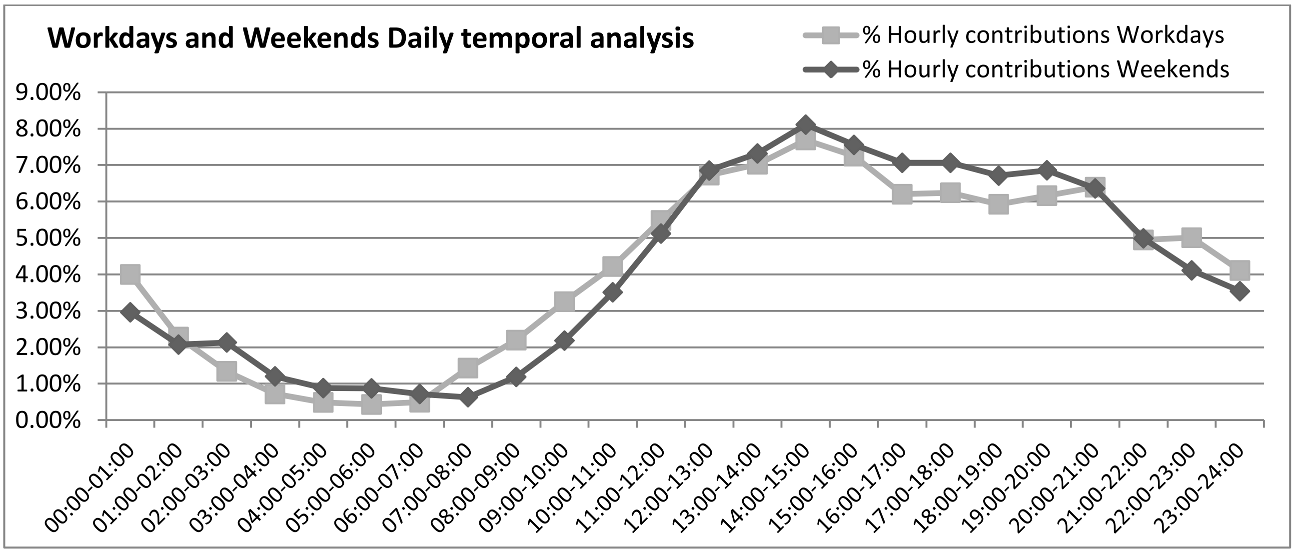


Figure 33. Daily SMGI percentages for workdays and weekends.

The obtained results provide information about the SMGI temporal distribution for different time periods; however, in order to obtain a further level of details, the analysis deals with the investigation of temporal patterns coupled with the spatial distribution of SMGI. Hence, the analysis of SMGI spatial distribution related to seasons and day periods is respectively conducted. The analysis of spatial distribution is conducted by means of ESRI ArcMap Kernel Density tool, which calculates a magnitude per unit area from the point features, relying on a kernel function in order to adjust a smoothly tapered surface to each point in the SMGI dataset. First of all, the analysis concerns the seasonal spatial distribution of SMGI. The results are shown in next figures (Fig. 34, Fig. 35, Fig. 36, Fig. 37), stressing the area presenting the highest density by means of a natural breaks (Jenks) classification based on 8 classes. The natural breaks classes aim to maximize the differences between classes exploiting the natural groupings inherent in the dataset.

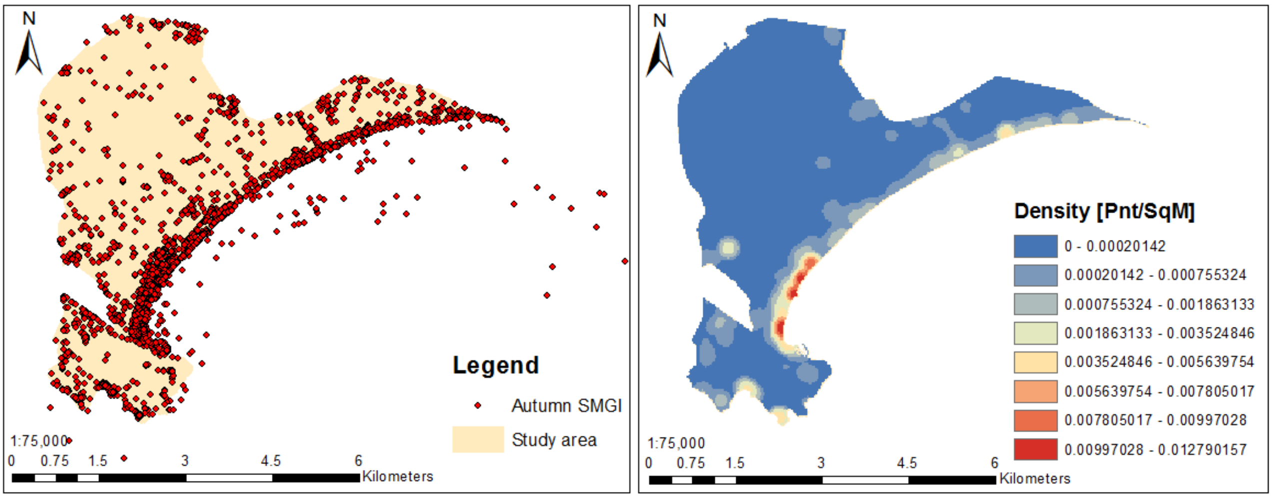


Figure 34. Autumn SMGI and resulting Kernel Density.

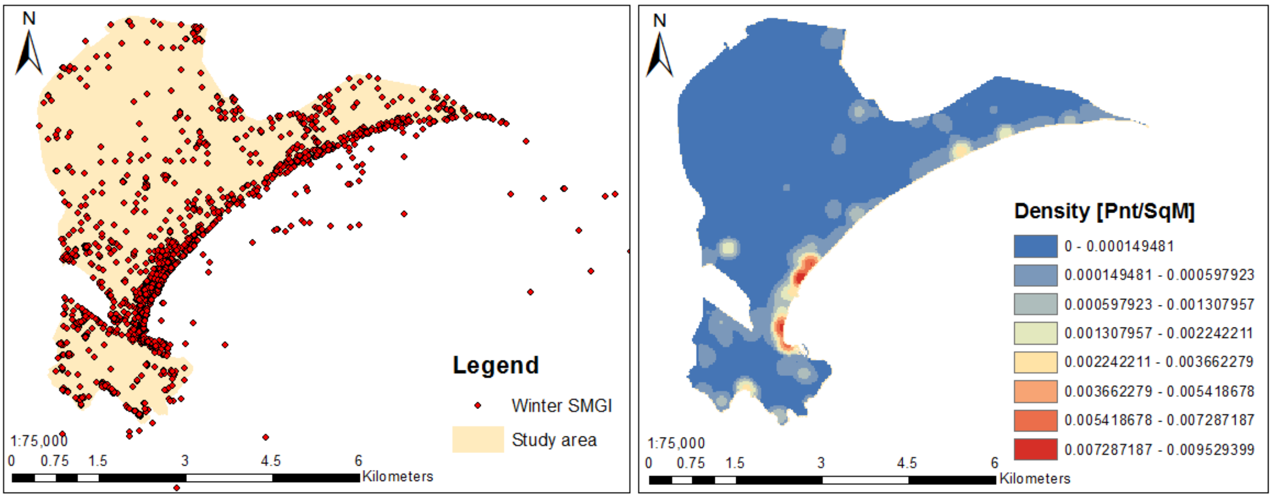


Figure 35. Winter SMGI and resulting Kernel Density.

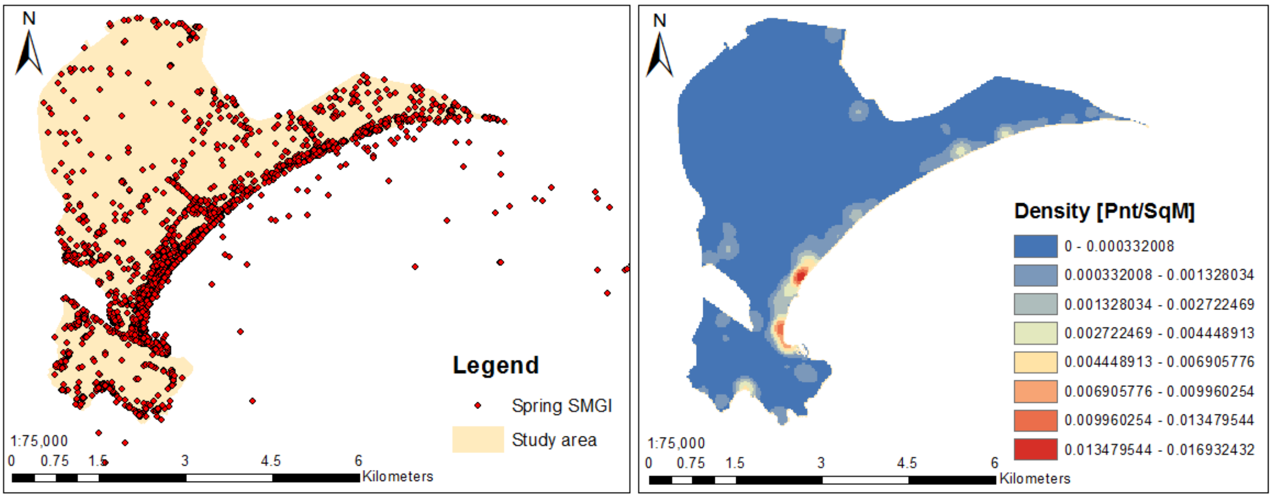


Figure 36. Spring SMGI and resulting Kernel Density.

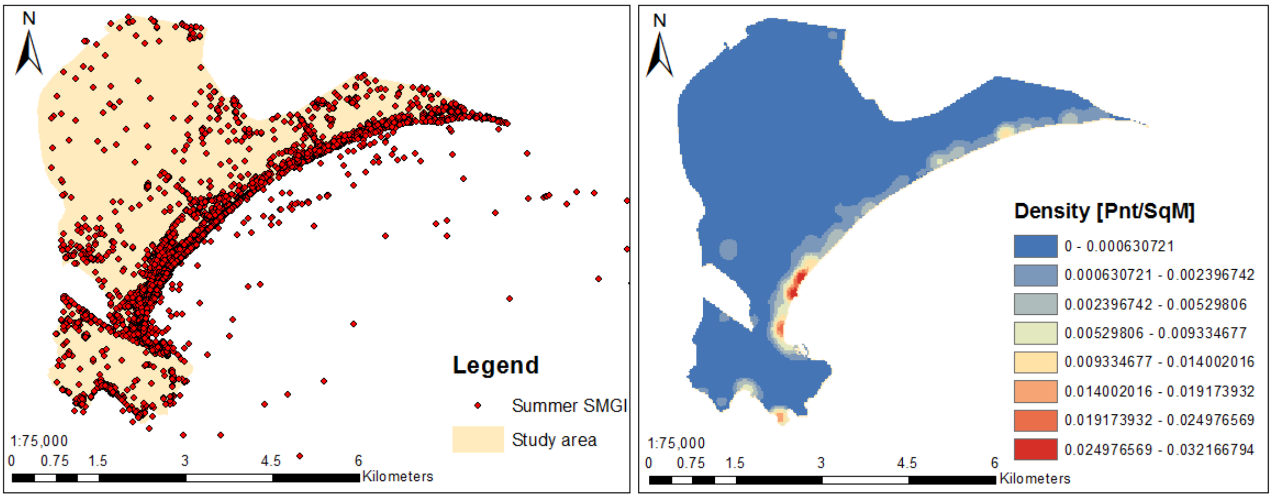


Figure 37. Summer SMGI and resulting Kernel Density.

The Kernel density results for the different seasons show homogeneous results among the SMGI spatial distribution. As a matter of fact, the areas exposing the highest density are similarly located in the first section of the Poetto Beach, bordering the touristic harbor. Nevertheless, a number of inner areas show major attractiveness during autumn and winter rather than during spring and summer. At the same time, the coastal area shows the major SMGI density. These slight differences are further following investigated in the chapter by means of local scale analyses.

Following the season spatial distribution investigation, the analysis concerns the study of daily distribution of SMGI. In this case the Kernel Density is used to compute the spatial distribution during morning (06:00 - 11.59), afternoon (12:00 - 17:59), evening (18:00 - 20:59) and night (21:00 - 05:59). The results are provided in next figures (Fig. 38, Fig. 39, Fig. 40, Fig. 41). Even in this case, the areas' density is represented through 8 classes of natural breaks classification.

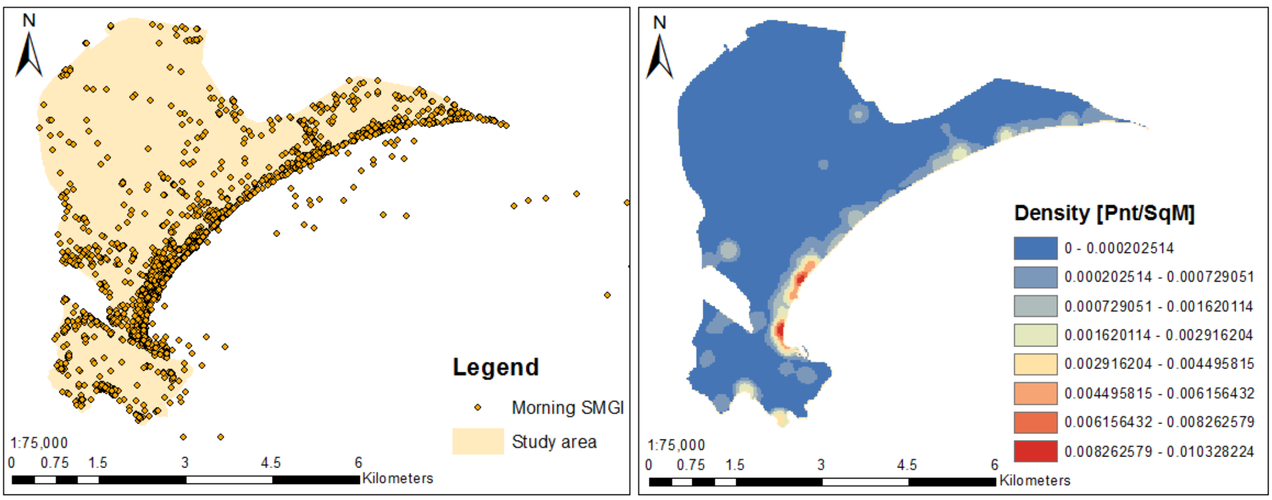


Figure 38. Morning SMGI and resulting Kernel Density.

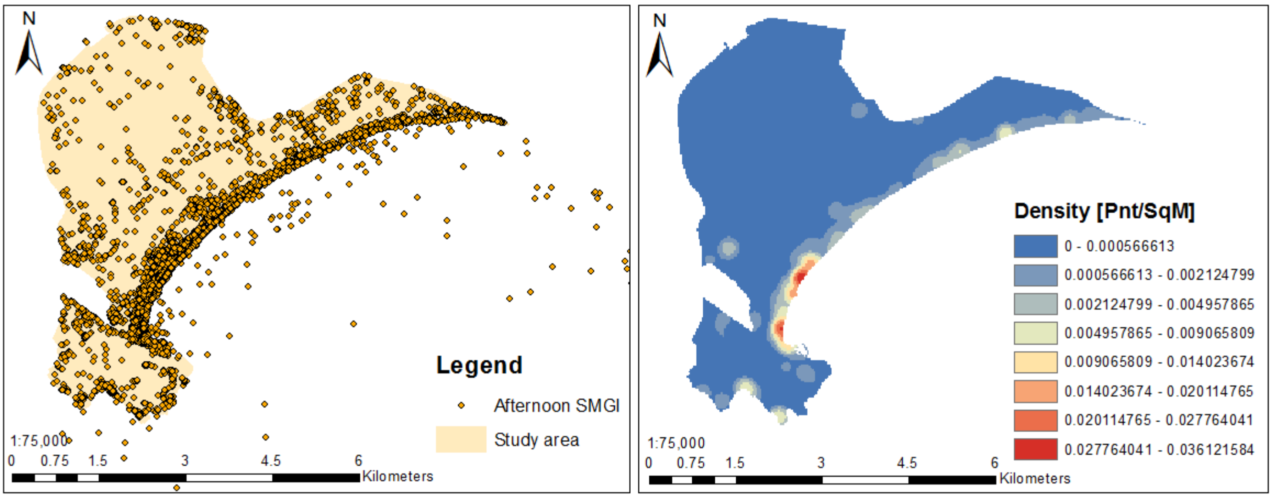


Figure 39. Afternoon SMGI and resulting Kernel Density.

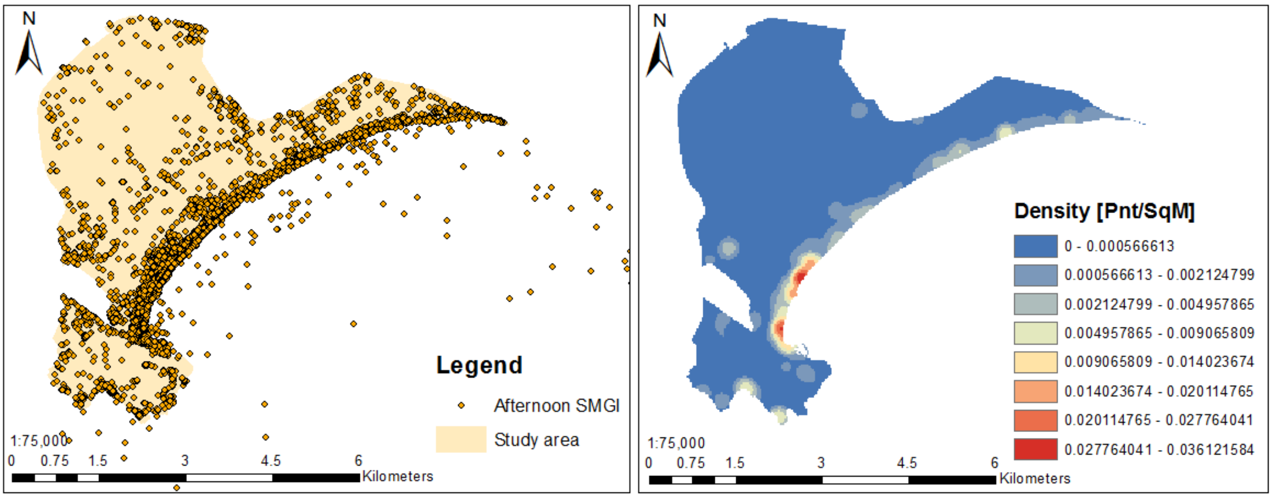


Figure 40. Evening SMGI and resulting Kernel Density.

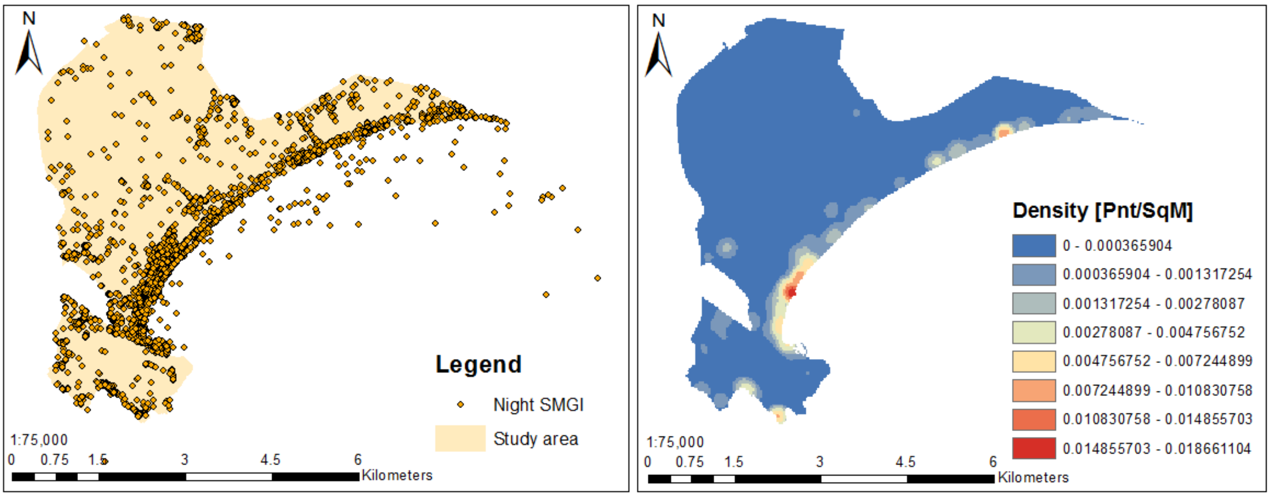


Figure 41. Night SMGI and resulting Kernel Density.

In this case, the analysis results show a SMGI homogenous spatial distribution during the different periods of the day. Notwithstanding the similarities, the night's density of contributions is more concerned on specific locations that are further investigated in the chapter by focused local scale analyses.

7.4 Spatial-temporal-user cluster analyses

The results of spatial and temporal investigations leads toward the development of further analyses in order to properly investigate the study area's geography and the underlying users dynamics. In particular, the higher density of SMGI in several locations requires the development of analytical methods able to identify, to classify and to interpret the reasons behind the users' interests toward these spaces. For this purpose, the cluster methodology applied on the Iglesias case study, is again conducted in this case study. Hence, the DB-SCAN algorithm and its slightly modified version FB-DBSCAN are used to compute clusters based on the spatial density of points for the first SMGI dataset and the second one, respectively. The cluster analysis on the first SMGI dataset is computed in order to identify the areas attracting the major interest of involved users, meanwhile the second SMGI dataset clustering is performed for detecting the potential residential locations of users, enabling the following user profiling. Indeed, as earlier demonstrated, many users shared a limited number of contributions in the study area both in space and time during the extraction period, and this circumstance raises the opportunity to consider such users as outer residents from Sardinia. Nevertheless, this hypothesis requires to be evaluated, hence the whole users' contribution for the extraction period, provided by the second SMGI dataset, is considered in order to detect outlier residential locations. The analysis on the first SMGI dataset related to the study area is conducted by means of the DB-SCAN algorithm that requires exclusively the threshold distance ε (*epsilon)* and the number of points to define a cluster (*min_pts*). However, there is again no opportunity to set the preferable values of the parameters before the computation, and furthermore in this case, the clusters are mainly related to a specific public space, requiring the evaluation of the results of different parameters' combinations. In order to obtain a feasible clustering, the analysis is conducted iteratively using the ε and *min_pts* values showed in Table 10.

DB-SCAN Parameters	
Threshold distance ε [meters]	**Minimum number of points (min_pts) [pts]**
50	5
50	10
50	15
15	10
15	15
10	10
10	15
5	10
5	15

Table 10. DB-SCAN parameters used to compute clusters in the study area.

In order to evaluate the differences between the parameters combinations, the analysis starts from the less restrictive pair (ε = 50 and min_pts = 5) until the most restrictive one (ε = 5 and min_pts = 15). In this case, the notable difference between parameters value used during the iterations is required in order to address the uncertainty related to the physical dimension of the locations belonging to the public space, which may range from a few square meters, in the case of a specific local, to notable dimensions, such as for instance in the case of a part of the Poetto beach or of the Molentargius Park. Indeed, a large threshold distance and a small number of included points may ease the identification of extensively distributed clusters, while small threshold distance and a substantial number of points identify highly frequented but localized clusters.

The assessment of the cluster analysis' results allow the identification of the following parameters' pair, which proved to be the most suitable for the purpose of the study: ε = 5 meters and min_pts = 10. As a matter of fact, the results demonstrate that the higher ε threshold distances lead towards the identification of clusters with excessive dimensions, usually joining different locations, thus limiting the opportunities to discriminate among bordering locations, which may attract differently the users interest. On the other hand, further increasing the *min_pts* value prevents the detection of several clusters, due to limited threshold distance and to the lower attractiveness of several locals or locations, which expose a modest participation of users. In addition, decreasing the *min_pts* value provoke the identification of a sheer number of clusters, often concerning exclusively the interest of a single user for a limited time period, causing false positives in clusters detection. The selected values are able to cover the dimension of small-medium public places, as well as, to detect popular locations in the study area.

The clustering analysis with the above parameters values enables the identification of 220 clusters within the study area, comprising more than 22K photos. The results demonstrate the major interest of users toward the coastal area where several clusters are identified. In particular the first section of the Poetto beach, near the 'Marina Piccola' touristic harbor, exposes the highest presence of clusters, probably related to the presence of many popular places among users. However, the clusters are present across the whole beach till the Quartu Sant'Elena municipality.

Conversely, in the 'Molentargius' Regional Park area the clusters are extremely limited, denoting a modest interest from users or, more probably, the habit to share sparsely distributed photos in the area, which arises hurdles in identifying clusters through the SMGI density. The detected clusters and their belonging SMGI contributions are shown in Figure 42.

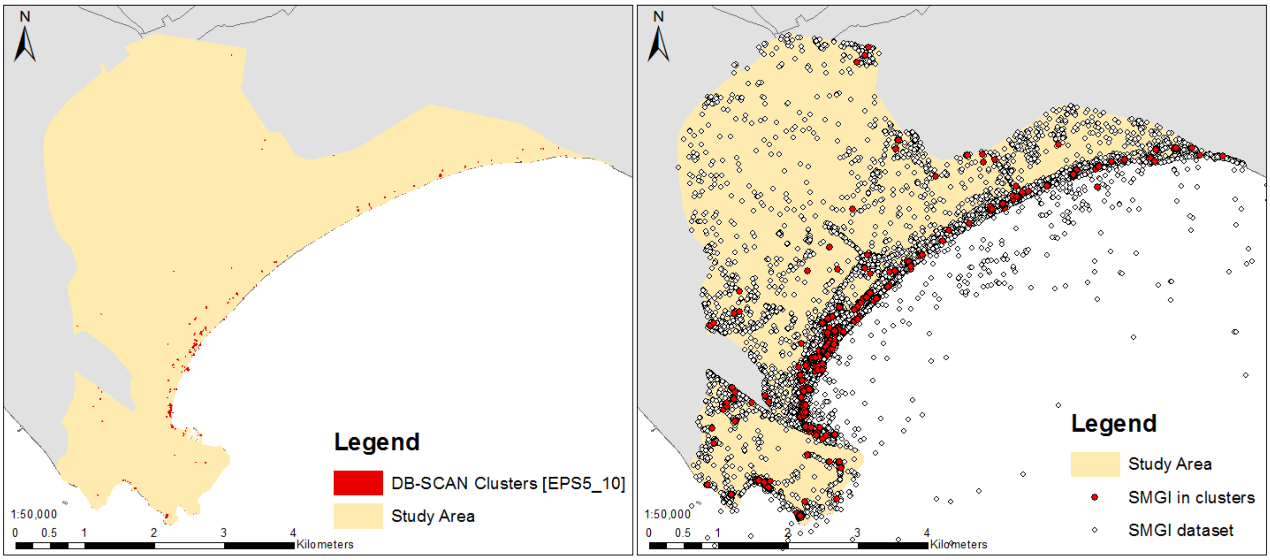


Figure 42. DB-SCAN resulting clusters (left) and SMGI belonging to clusters (right).

A closer look to the resulting clusters, as shown in Figure 43, demonstrates how the used parameters (ε = 5 meters; *min_pts* = 10) are able to discriminate among the locations although their spatial proximity.

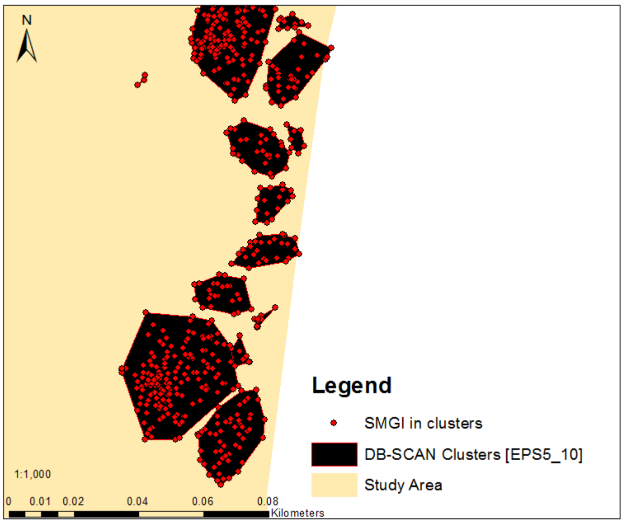


Figure 43. DB-SCAN resulting clusters and in-cluster SMGI (zoomed).

After the first cluster analysis, which was exclusively conducted on the SMGI related to the study area, the second cluster analysis deals with the investigation of potential users residential locations. Hence, the Feature-Based DB-SCAN is used on the second SMGI dataset, concerning all the users' contributions during the extraction period, to compute the clusters attracting the major interest of each individual user. Indeed, the FB-DBscan tool processes iteratively the SMGI dataset after performing a selection for each value in a specific attribute, in this case the user. Nevertheless, the threshold distance (ε) and the minimum number of points in each cluster (*min_pts*) have to be selected in order to obtain a feasible result for the identification of residential locations. Taking advantage of the previous clustering analysis results obtained from the Iglesias case study, the computation is carried out for two different combinations of parameters, which are provided in Table 11. As a matter of fact, these values appear to be the most suitable to obtain a proper clustering of the dataset, enabling the identification of small-medium sized clusters that are able to cover the dimension of a medium-size fabric.

FB-DBSCAN Parameters	
Threshold distance ε [meters]	**Minimum number of points (min_pts) [pts]**
10	10
15	10

Table 11. FB-DBSCAN parameters used to compute clusters in users SMGI dataset.

Firstly, the clustering analysis is conducted with the parameters ε = 10 meters and *min_pts* = 10 points, resulting in 6263 clusters that belong to 3711 different users. Secondly, the clustering analysis through FB-DBSCAN is carried out with the parameters ε = 15 meters and *min_pts* = 10 points and results in 7013 clusters, which belong to 4086 different users. In the light of the obtained results, the second clustering option is selected due to the major number of involved users, and the results are shown in Figure 44, where each cluster is represented by its centroid.

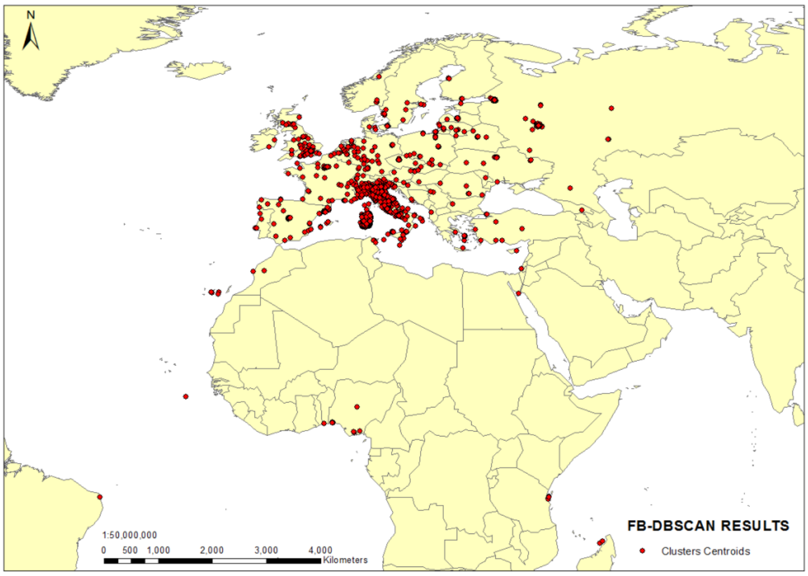


Figure 44. FB-DBSCAN clusters.

At this point, concluded the FB-DBSCAN analysis, the methodology concerns the classification of the resulting users clusters. According to the methodology applied in the Iglesias case study, the clusters are classified relying upon several parameters related to both the cluster’s features and the included contributions.

However, in this case most of the clusters to be classified are located outside the Sardinia Region, dismissing the opportunity to evaluate the overlap between the cluster’s centroid and an official building footprint. Therefore, in order to classify a cluster as residential, two different combinations of the parameters are used, as listed in next tables (Table 12 and Table 13).

COMBINATION 1			
CONDITION: Cluster's centroid within an official building footprint or within a 1 meter gap			
Parameters	***Description***	***Units of measure***	***Value***
Cluster Area	The cluster area	Square meters	=> 20
Time Span Among Photos	The time passed between the first contribution and the last one in the cluster	Days	=> 180

Table 12. Parameters for clusters classification: combination 1

COMBINATION 2			
CONDITION: Cluster's centroid outside an official building footprint or not evaluable			
Parameters	***Description***	***Units of measure***	***Value***
Cluster Area	The cluster area	Square meters	=> 10
Number of Contributions	The total number of contributions comprised in the cluster	Number of contributions	=> 15
Cluster Density	The ratio between the cluster's area and the number of contained contributions	Square meters	=> 4
Time Span Among Photos	The time passed between the first contribution and the last one in the cluster	Days	=> 180

Table 13. Parameters for clusters classification: combination 2

In addition, a number of users may expose multiple personal clusters, which are able to satisfy both the parameters combinations. In order to avoid this issue, the results are parsed by means of an ad-hoc designed and developed tool of SPATEXT, which processes the results and selects exclusively one residential cluster for each user. The tool confirms as residential the cluster resulting from the first parameters' combination, which exploits the cluster overlap with an official building footprint, and dismisses the other users' clusters if existent. However, if multiple users' clusters result from the second parameters combination, the tool selects as residential the user cluster that exposes the major number of contained contributions. Eventually, in the case of the same contained SMGI, the cluster with the major time span among contributions is selected.

The classification of clusters, performed by means of the aforementioned parameters and conditions, leads to the identification of 2557 residential clusters belonging to users who visited the study area during the extraction period. In addition, the classification results are assessed by means of a manual visual assessment through satellite imagery on a sample of 100 hundred randomly selected clusters' centroids, which confirms the residential building identification. Among the identified residential locations, 1975 are located in Sardinia, 375 in Italy, while 203 are sparsely distributed worldwide. The results of residential clusters' classification are shown in Figure 45, where each identified cluster is represented by its centroid.

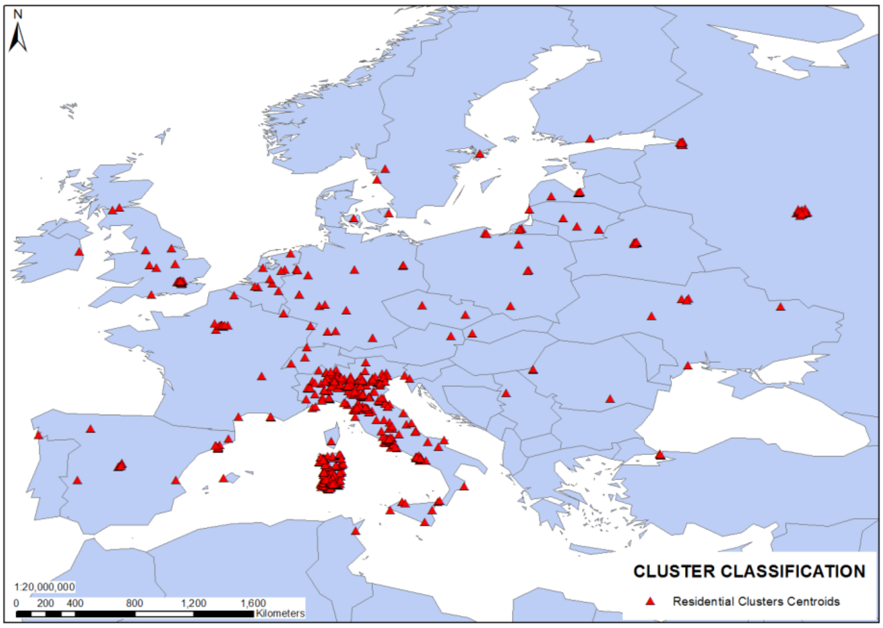


Figure 45. FB-DBSCAN Clusters classification results.

The obtained results represent a fundamental step to conduct the following user profiling analysis, inasmuch the user residential locations, placed within the Sardinia territory, may be classified thanks to the development of the geodemographic classification of the Region census tracts. This way, the resulting key characteristics of each classified census tract, wherein a residential cluster is located, may be assigned to the respective user profile.

7.5 Geodemographic classification

The geodemographic classification of the Sardinia Region is conducted using solely the official census data of the year 2001, provided by the ISTAT for the Sardinia Region. In spite of more recently updated information, hailing from the 2011 national census, the geographic coverage of this data is not feasible for the purpose of the study. In addition, secondary unofficial sources of information are not used in the analysis in order to avoid potential issues related to different geographic scales, accuracy, as well as reliability of data.

The methodological approach used for developing the Sardinian geodemographic classification is based upon the methodology adopted by Vickers and Rees (2007) for the Output Area Classification (OAC) of the United Kingdom. Therefore the methodology consists of the following 4 steps:

1. clustering elements identification;
2. variables selection;
3. clustering analysis;
4. clusters profiling.

From an analytical perspective, each identified step may require different stages, which have to be completed in order to accomplish a proper geodemographic classification of the Sardinia territory.

7.5.1 Clustering elements identification

The first step of the geodemographic classification concerns the identification of clustering elements, namely the objects to cluster. These elements should define a complete geographical coverage of the study area and have to be representative of the resulting cluster structure to achieve. Hence, in order to carry out a feasible geodemographic classification and to ensure a complete geographical coverage, the 13,325 census tracts related to the Sardinia Region for the year 2001, provided by the Italian official census data, are used. The 2001 census tracts are able to cover the whole Region and are enriched with the set of socio-economic variables, collected for the census and reliable in terms of data collection methodology, which may enable the geodemographic classification.

7.5.2 Variables selection

The selection of variables is a fundamental step for any clustering analysis, inasmuch the variables represent the measurements taken on each entity or area for enabling clustering (Milligan, 1996). In order to develop the geodemographic classification, the variables should be significant, that is to be able to discriminate among the socio-economic characteristics of the Sardinia population. By reviewing the official 2001 census data, available at the census tract scale, 199 different variables are identified. However, the variables should be included solely if able to further discriminate among clusters and irrelevant variables should be avoided in order to ease the pattern recognition. In this respect, the available census variables are classified and evaluated for their inclusion in the classification, leading toward the identification of 5 main categories:

- demographics;
- household composition;
- housing typology;
- socio-economic status;
- employment condition.

Moreover, the census variables evaluation confirms the presence of many redundant or irrelevant variables, which are not selected for the analysis. The variables are selected by considering their constancy across the whole census tracts, by avoiding potential uninteresting geographic distributions and by fostering the development of composite variables. Indeed, composite variables may be created by merging the values of two related variables, which expose similar patterns and share the same denominator. The composite variables may be used in order to avoid the presence of highly correlated variables or to depict characteristics concerning exclusively a small group of population (Vickers and Rees, 2007). In addition, the evaluation assesses the uncertainty in measurements, the temporal consistency and the opportunity to standardize either all variables or exclusively the variables affected by age distribution to a common scale. Therefore, a number of variables mainly concerning the housing typology are standardized by range method, meanwhile several variables related in particular to demographics are coupled in order to define an unique composite variable. All the selected variables measurements not obtained by the range method normalization are expressed in percentage in order to avoid issues due to excessively small or big values, which may affect the geodemographic classification results. Finally, a dataset consisting of 43 variables, originated from the assessment of the 199 initial ones, is obtained from the evaluation stage. The selected variables are listed in Table 14.

VARIABLE	DESCRIPTION	CATEGORY
V1	Percentage of resident population aged 0-4 years	Demographics
V2	Percentage of resident population aged 5-14 years	Demographics
V3	Percentage of resident population aged 25-44 years	Demographics
V4	Percentage of resident population aged 45-64 years	Demographics
V5	Percentage of resident population aged > 65 years	Demographics
V6	Percentage of resident population Afrikans	Demographics
V7	Percentage of resident population Asian	Demographics
V8	Percentage of resident population Caucasic of Hispanic	Demographics
V9	Density Pop/sqKm [normalized by range method]	Demographics
V10	Percentage of unmarried	Household composition
V11	Percentage of married	Household composition
V12	Percentage of separated, divorced or widowed	Household composition
V13	Percentage of single person household	Household composition
V14	Percentage of couple household	Household composition
V15	Percentage of household of 3-4 persons	Household composition
V16	Percentage of household > 5 persons	Household composition
V17	Percentage of house owners	Housing Typology
V18	Percentage of house tenants	Housing Typology
V19	Percentage of house with heating	Housing Typology
V20	Average rooms per house [normalized by range method]	Housing Typology
V21	Average persons per room [normalized by range method]	Housing Typology
V22	Atypical houses [normalized by range method]	Housing Typology
V23	Average area per house [normalized by range method]	Housing Typology
V24	Percentage of old houses 1919 - 1971	Housing Typology
V25	Percentage of recent houses > 1972	Housing Typology
V26	Percentage of building with 1 or 2 roofs	Housing Typology
V27	Percentage of condominium	Housing Typology
V28	Percentage of building with > 3 apartment numbers	Housing Typology
V29	Percentage of people with High education level	Socio economic
V30	Percentage of people with Medium education level	Socio economic
V31	Percentage of people with Low or No education level	Socio economic
V32	Percentage of commuters	Socio economic
V33	Percentage of inner-municipality commuters	Socio economic
V34	Percentage of employed	Socio economic
V35	Percentage of Students	Employment condition

V36	Percentage of Not employed (retired or other condition)	Employment condition
V37	Percentage of businessman of freelancer	Employment condition
V38	Percentage of salaried worker	Employment condition
V39	Percentage of employers in Agriculture	Employment condition
V40	Percentage of employers in Industry	Employment condition
V41	Percentage of employers in Public Services	Employment condition
V42	Percentage of employers in Trade, Restaurant, Transport, Communication	Employment condition
V43	Percentage of employers in Financial intermediation and business	Employment condition

Table 14. SMGI Analytics Geodemographics: selected socio-economic variables.

The listed variables demonstrate that a large number of early variables were rejected or merged for producing the final dataset. The original census variables offer a detailed overview of the socio-economic characteristics of the population, as well as a clear picture of the household composition. housing typologies and employment conditions.

Nevertheless, most of this information may be redundant or not feasible for the geodemographic classification. As a matter of fact, the variables related explicitly to the population gender are rejected inasmuch gender is not informative for the study purpose. Similarly, several variables related to population's age and household composition are merged due to their highly correlation. Moreover, many of the selected variables, concerning the housing typology and the employment conditions, are the result of a merging activity developed in order to avoid the presence of high intra-category variables correlation.

After the variables' selection step, the resulting dataset is further assessed in order to evaluate the potential presence of variables' pairs exposing strong intra-category and inter-categories correlation. This stage is performed by means of the Pearson correlation coefficient (Pearson, 1895), which is a measure of the linear correlation between two variables. Pearson's correlation coefficient is calculated as the covariance of the two variables divided by the product of their standard deviations.

$$\rho_{X,Y} = \frac{cov(X,Y)}{\sigma_X \sigma_Y} \qquad where\ cov = covariance,\ \ \sigma_X\ and\ \sigma_Y = standard\ deviation$$

The resulting value is comprised between +1 and -1, wherein 1 is the total positive correlation, 0 represents no correlation between variables and -1 is the total negative correlation. The Pearson correlation coefficient is calculated for each pair of variables, and the p-value, representing the probability of obtaining a significant result is at the same time calculated. A small p-value means a large significance of the obtained results. The results of correlation analysis between variables' pairs are provided in attachment to the thesis.

In spite of some correlations found in the variables, the dataset is not modified. As a matter of fact, the presence of highly correlated variables' pairs in the dataset may be used to better discriminate among groups during the geodemographic classification due to their notable descriptive and predictive power

(Voas and Williamson, 2001; Vickers and Rees, 2007). In the light of these considerations, all the 43 variables are confirmed and included in the dataset for the geodemographic classification.

Finally, the last stage deals with the checking of the census tracts dataset, in order to evaluate the constancy and reliability of the measurements of the variables, selected among all the considered entities. The dataset assessment identifies the existence of several census tracts not populated, which may considerably affect the geodemographic classification, for the presence of zero values in variables measurements. Therefore, the census tracts without resident population are removed from the dataset and automatically classified as "No Population". This operation decreases the number of entities to classify from the original 13,325 census tracts to the 9,488 inhabited ones, which are processed by means of the clustering analysis.

7.5.3 Clustering analysis

The geodemographic classification of the Sardinia territory aims to define a two-tier hierarchical classification able to achieve a synthetic description of the different population groups relying on the identified variables. The first hierarchical level may provide a general description of the population groups meanwhile the second hierarchical level may depict the particular features of each identified sub-group.

The clustering method, selected for developing the clustering analysis on the Sardinia census tracts, is an optimized version of the K-Means algorithm, namely the K-Means++, which is implemented by the '*grouping analysis'* tool of ESRI ArcGIS. The algorithm partitions the n entities of the input dataset, in this case the census tracts, into an established k number of clusters, by assigning each entity to the cluster exposing the nearest mean. Each entity is represented as a 43-dimension vector, where each dimension is represented by the selected variable measurement. Each census tract is automatically assigned by the tool to the cluster presenting the least within-cluster sum of squares, which is the squared Euclidean distance of the 43 selected variables. The tool runs iteratively for a specified number of times or until the final convergence, namely when no moves of entities occur during a complete iteration.

One of the main problems arising to conduct a successful clustering analysis lies in selecting the correct number of clusters to partition the entities, in this case the census tracts. According to the approach adopted by Vickers and Rees (2007), the most feasible solution is to run iteratively the clustering analysis varying the number of clusters to find and evaluating the obtained results in terms of entities belonging to each cluster and pseudo F-Statistics, namely the ratio between intra-cluster variance and inter-cluster variance (Calinski and Harabasz, 1974). In particular, the clusters' size performs a central role in geographic classification, inasmuch clusters should be as closely sized as possible to each other. Indeed, unevenly sized clusters, overall in the first hierarchical level, may cause issues in the next hierarchical level segmentation.

The ‘*grouping analysis’* tool of ESRI ArcGIS is run iteratively for different values of K, namely the number of clusters or groups for partitioning the census tracts of Sardinia. According to the suggestions of Callingham (2003), the first hierarchical level of a geodemographic classification should consist of about 6 groups. Therefore, the grouping analysis is used to process the dataset for a K value ranging from 4 to 8 and the results are evaluated in order to detect the most suitable solution both in terms of Pseudo F-Statistics and number of entities belonging to each group. The ‘*grouping analysis’* tool automatically calculates the mean, minimum, maximum and median Pseudo F-Statistics values, enabling a direct evaluation of the resulting clusters. The results are provided in Table 15 and shown in figure 46, 47, 48 and 49, respectively.

GROUPS	**Pseudo F-Stats Mean**	**Pseudo F-Stats Min**	**Pseudo F-Stats Max**	**Pseudo F-Stats Median**
4 Groups	593.1204	540.6181	628.9119	607.3049
5 Groups	539.7334	455.2414	570.6642	540.3529
6 Groups	508.1854	474.6544	525.4685	509.4646
7 Groups	469.2808	437.4045	500.8139	471.0736
8 Groups	436.9851	415.1371	452.9164	439.3360

Table 15. Pseudo F-Statistics results.

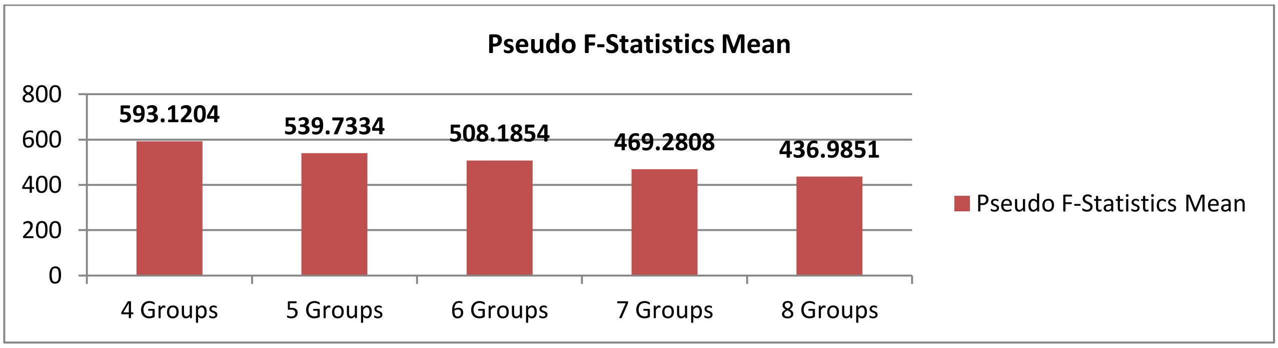


Figure 46. Pseudo F-Statistics Mean.

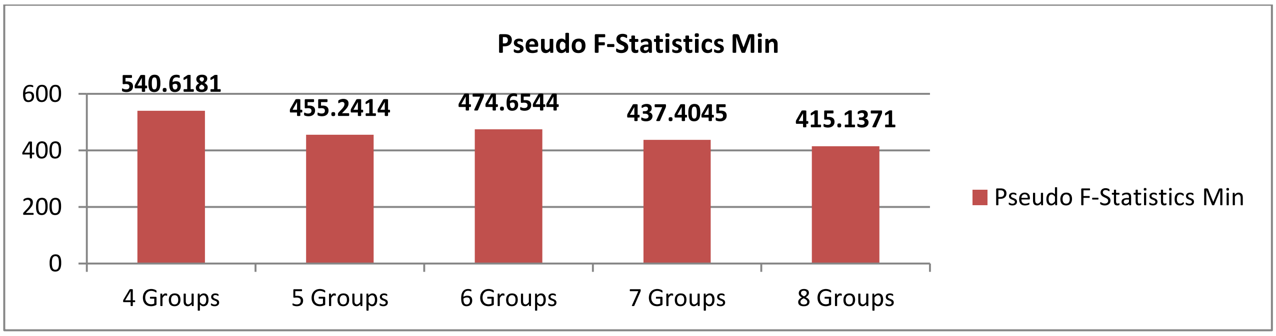


Figure 47. Pseudo F-Statistics Minimum.

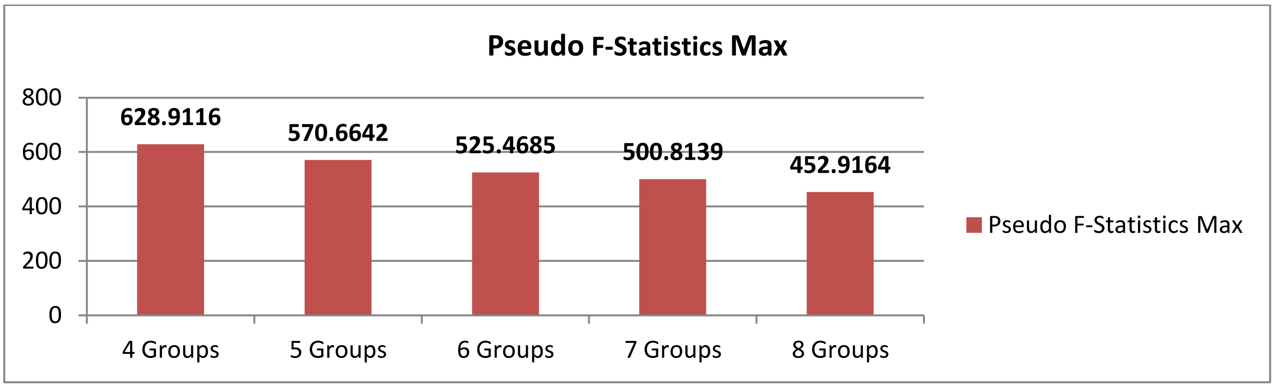


Figure 48. Pseudo F-Statistics Maximum.

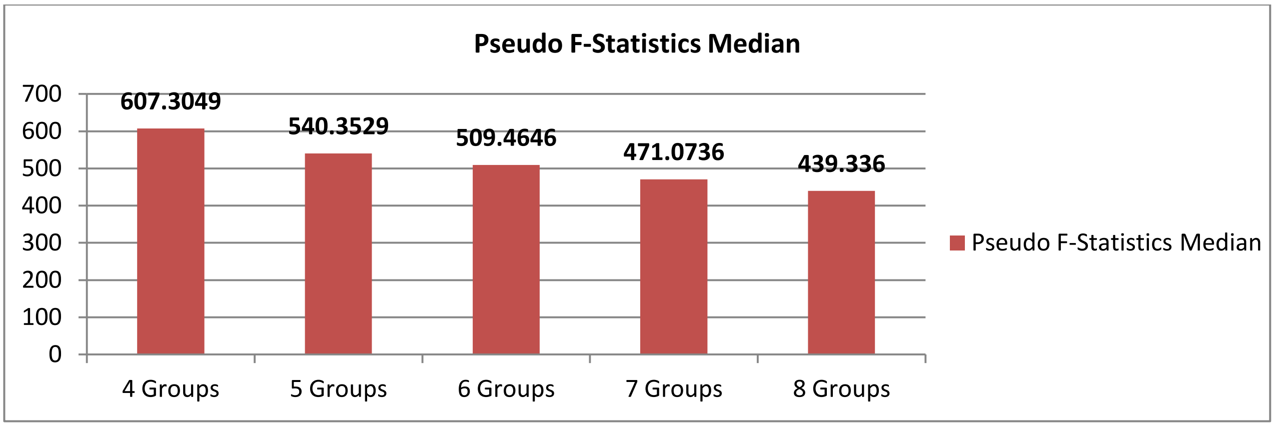


Figure 49. Pseudo F-Statistics Median.

The obtained results measure the grouping effectiveness using the Pseudo F-Statistics, namely a ratio reflecting the intra-cluster similarity and the inter-cluster difference. The largest F-Statistics value highlights the number of groups most effective at segmenting the entities according to the included variables' measurements. Therefore, the most feasible solution appears to be the census tracts dataset partitioning in 4 groups; however, this result has to be assessed in terms of elements belonging to each group. The assessment of census tracts assigned to each group is computed considering two parameters: 1) the maximum gap between the mean cluster size and the group containing the biggest or the smallest number of entities; 2) the maximum intra-group gap that is the difference between the group with the major number of elements and the group with the minimum one. The results of the assessment are provided in Table 16 and 17 and Figure 50 and 51.

	Entities belonging to each group							
K number	**1 group**	**2 group**	**3 group**	**4 group**	**5 group**	**6 group**	**7 group**	**8 group**
4	570	2418	5436	1064				
5	3201	3261	2105	265	656			
6	554	1817	2926	256	902	3033		

7	546	1084	265	1192	559	2852	2990	
8	574	1728	15	234	254	2979	870	2834

Table 16. Census tracts included in each group for different K values.

	Assessment Parameters				
K number	**Mean**	**Max value**	**Min value**	**Max Gap from Mean**	**Max intra-cluster Gap**
4	2372	5436	570	|3064|	4866
5	1897.6	3261	265	|1363.4|	2996
6	1581.33	3033	256	|1451.67|	2777
7	1355.43	2990	265	|1634.57|	2725
8	1186	2979	15	|1793|	2964

Table 17. Results of assessment parameters used to evaluate clustering results.

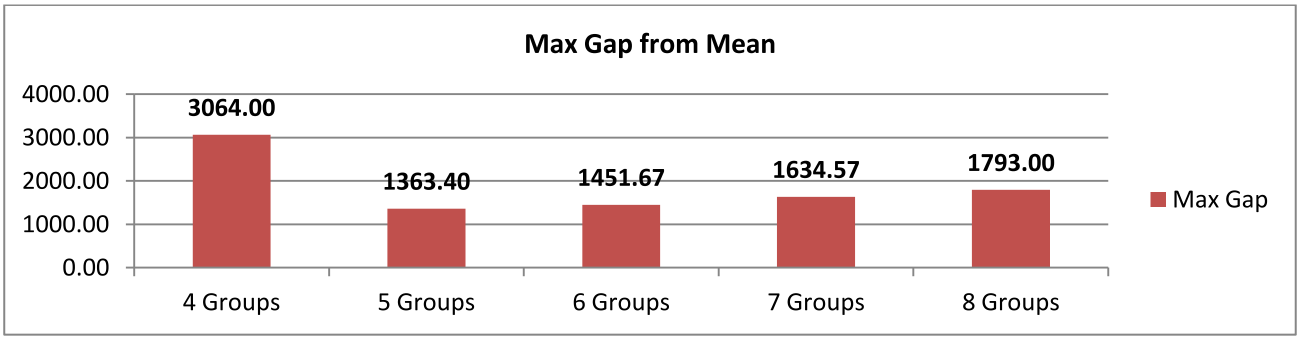


Figure 50. Max gap from Mean parameter results.

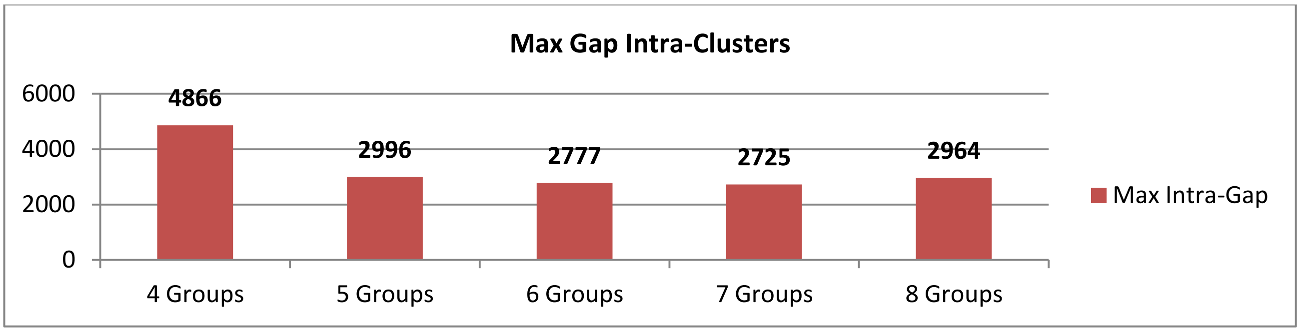


Figure 51. Max gap intra-clusters parameter results.

The assessment results demonstrate that the entities segmentation in groups is strongly heterogeneous and may change significantly according to the selected K value. Moreover, the partitioning solution in 4 groups, suggested by the Pseudo F-Statistics, appears not feasible due to the excessive gaps in clusters size. Therefore, the most suitable solution to define the number of K groups is selected comparing the obtained results for each group, in order to satisfy both the grouping effectiveness and the clusters size requirements. The Table 18 shows the comparison among the different partitioning solutions for the first hierarchical level of the geodemographic classification.

	Clustering Parameters		
K number	**Pseudo F-statistics Mean**	**Max Gap from Mean**	**Max Gap intra-cluster**
4	593.1204	3064	4866
5	539.7334	1363.40	2996
6	508.1854	1451.67	2777
7	469.2808	1634.57	2725
8	436.9851	1793.00	2964

Table 18. Comparison among K-Means clustering solutions ranging from 4 to 8 groups.

The K=4 solution offers the most suitable Pseudo F-Statistics value, but at the same time, it leads to the creation of excessively unevenly sized clusters as shown from the clusters size assessment. Similarly, the solutions K=7 and K=8 are not feasible both in terms of low Pseudo F-Statistics coefficients and clusters sizes. The most suitable solutions are represented by K=5 and K=6, which provide both satisfactory Pseudo F-Statistic values and acceptable differences among clusters sizes. Nevertheless, the K=6 solution is preferred due to the lowest intra-cluster gap and to the Pseudo F-Statistics' mean compared to the K=5 solution, as shown in Figure 52.

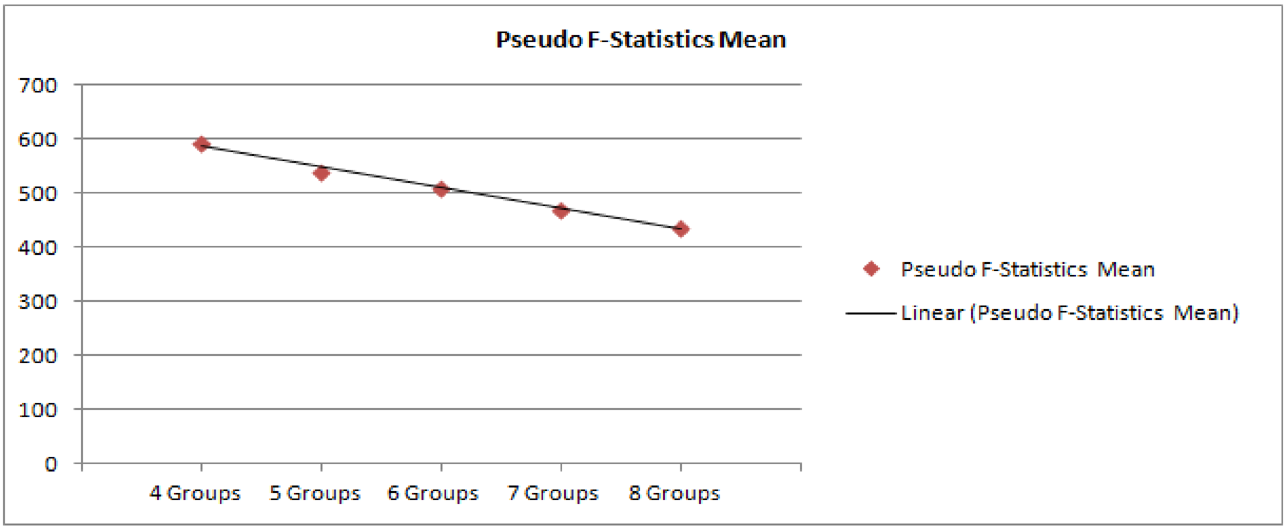


Figure 52. Pseudo F-Statistics Mean:

In the light of the obtained findings, the clustering analysis is conducted through the '*grouping analysis*' tool of ArcGIS partitioning the Sardinian census tracts in K=6 groups. The resulting groups are representative of different population segments, which are obtained processing the considered variables by maximizing the inter-groups differences and by minimizing the intra-group ones. The clustering result is provided in Figure 53, wherein both the census tracts belonging to each identified group and the census tracts without population are figured.

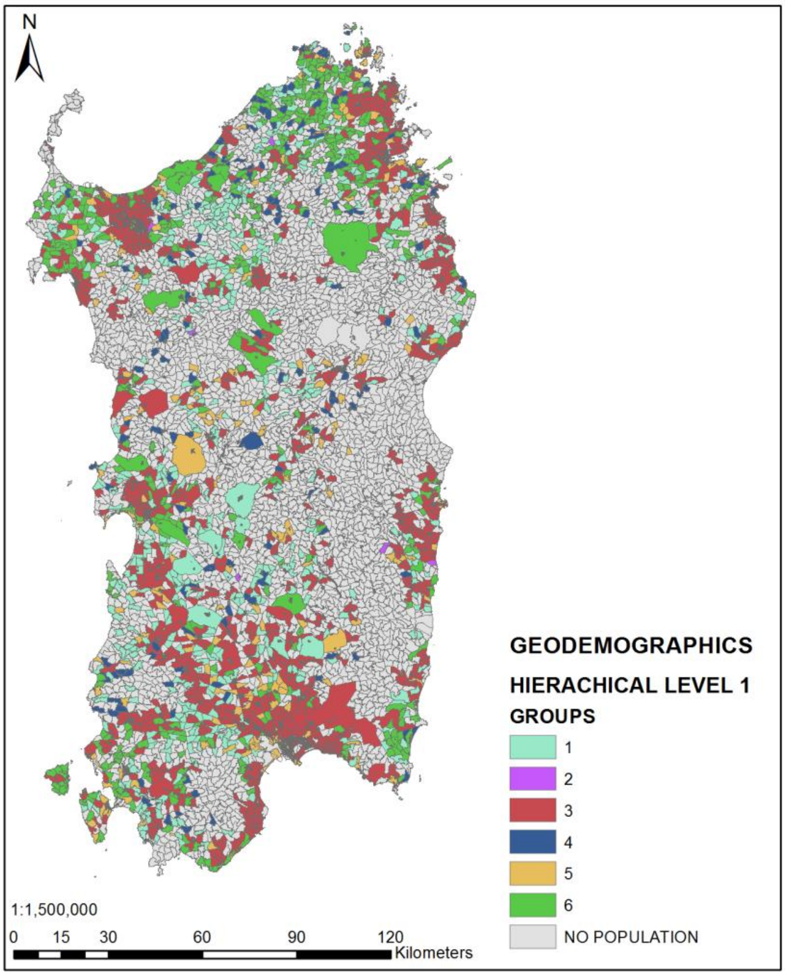


Figure 53. Geodemographic classification: first hierarchical level.

In addition, the analysis results foster the identification of several parameters related to the variables used for the dataset partitioning, as listed in Table 19. The larger the R^2 is for a particular variable, then that variable is more suitable for partitioning the dataset features, as provided by the variables V27 (Percentage of condominium), V26 (Percentage of buildings with 1 or 2 roofs), V39 (percentage of employers in agriculture) and V28 (Percentage of buildings with more than 3 apartments).

SUMMARY OF VARIABLES USED FOR GROUPING ANALYSIS					
VARIABLE	**MEAN**	**STD.DEVIATION**	**MINIMUM**	**MAXIMUM**	**R2**
V27	0.237316	0.317927	0	1	0.622944
V26	0.754042	0.324874	0	1	0.590787
V39	0.116978	0.219775	0	1	0.573559
V28	0.262574	0.342041	0	1	0.563438
V17	0.74342	0.250545	0	1	0.441517
V36	0.224966	0.158548	0	1	0.434108
V38	0.670715	0.25409	0	1	0.422352
V34	0.785628	0.202226	0	1	0.41375
V18	0.248359	0.242287	0	1	0.405269
V25	0.497904	0.371673	0	1	0.389028
V24	0.493453	0.371621	0	1	0.37919
V9	0.041842	0.055299	0	1	0.335233
V37	0.281044	0.226692	0	1	0.33263

V5	0.156347	0.144495	0	1	0.305208
V41	0.308335	0.21315	0	1	0.232063
V29	0.075605	0.108887	0	1	0.208095
V32	0.317934	0.18013	0	1	0.1825
V31	0.353972	0.187312	0	1	0.181696
V15	0.40777	0.215094	0	1	0.177733
V13	0.276376	0.224316	0	1	0.163644
V23	0.212974	0.059742	0	1	0.152708
V20	0.342628	0.077813	0	1	0.152589
V19	0.88521	0.188314	0	1	0.126417
V40	0.218146	0.187831	0	1	0.126202
V30	0.570423	0.170129	0	1	0.122922
V43	0.072401	0.101868	0	1	0.119487
V42	0.251468	0.19192	0	1	0.114981
V35	0.073589	0.063292	0	1	0.112022
V33	0.107391	0.117821	0	1	0.110274
V11	0.442916	0.150112	0	1	0.108087
V10	0.469353	0.145118	0	1	0.107496
V12	0.087731	0.098613	0	1	0.107268
V3	0.331623	0.148662	0	1	0.106759
V2	0.092847	0.076092	0	1	0.096264
V21	0.059369	0.028821	0	1	0.091381
V14	0.21151	0.15891	0	1	0.061306
V1	0.039597	0.047649	0	1	0.045026
V16	0.100549	0.129278	0	1	0.036829
V4	0.26142	0.146104	0	1	0.02558
V6	0.003986	0.033089	0	1	0.025449
V8	0.009539	0.059575	0	1	0.015994
V7	0.00172	0.021648	0	1	0.008442
V22	0.000245	0.010538	0	1	0.001406

Table 19. Summary of variables used for 'grouping analysis'.

7.5.3.1 Clustering analysis: second hierarchical level

After the conclusion of the first hierarchical level, the methodology concerns the creation of the second hierarchical level of the geodemographic classification. The adopted methodology, used for computing the second level, is based upon the aforementioned steps; however at this time, each first hierarchical level group is processed independently in order to find the most suitable number of sub-groups. According to Callingham (2003) and Vicker and Rees (2007), the second hierarchical level should consist of about 20 sub-groups. Since the first hierarchical level is composed by 6 groups, each of these groups should be partitioned in a number of sub-groups ranging from 2 to 4, in order to obtain 12 sub-groups at least or 24 groups at most. Therefore, the '*grouping analysis*' tool of ESRI ArcGIS is run iteratively for K values ranging from 2 to 4 on the features of each first level group. Then the results are assessed in terms of Pseudo F-Statistics and number of entities belonging to each sub-group in order to establish the preferable number of K groups for partitioning.

The clustering analysis of the first hierarchical level group 1 is conducted for K=2, K=3 and K=4, leading to the results shown in Figure 54, Table 20 and Table 21.

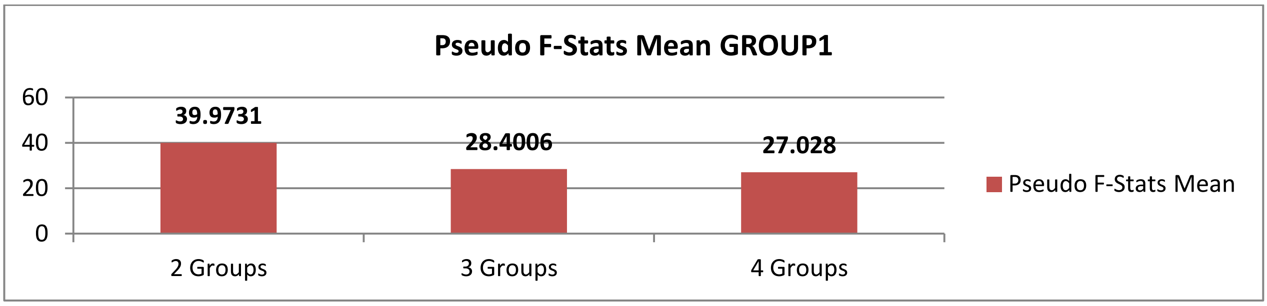


Figure 54. Pseudo F-Statistics Mean for partitioning the First Hierarchical Level Group 1 in sub-groups.

GROUP1	Entities belonging to each group			
K number	1 group	2 group	3 group	4 group
2	153	401	-	-
3	1	178	375	-
4	99	103	7	345

Table 20. Entities belonging to each sub-group of the First Hierarchical Level Group 1.

GROUP1	Assessment Parameters				
K number	Mean	Max value	Min value	Max Gap from Mean	Max intra-cluster Gap
2	277	401	153	|124|	4866
3	184.67	375	1	|190.34|	2996
4	138.5	345	7	|206.5|	2777

Table 21. Results of assessment parameters used to evaluate sub-groups clustering results.

The obtained results identify the solution K=2 as the most feasible for partitioning the first hierarchical level Group 1 features. As a matter of fact, this solution satisfies both the Pseudo F-Statistics and the clusters size assessment parameters. Similarly, the clustering analysis of the first hierarchical level Group 2 is conducted for K ranging from 2 to 4 and provides the results shown in Figure 55 and Table 22 and Table 23.

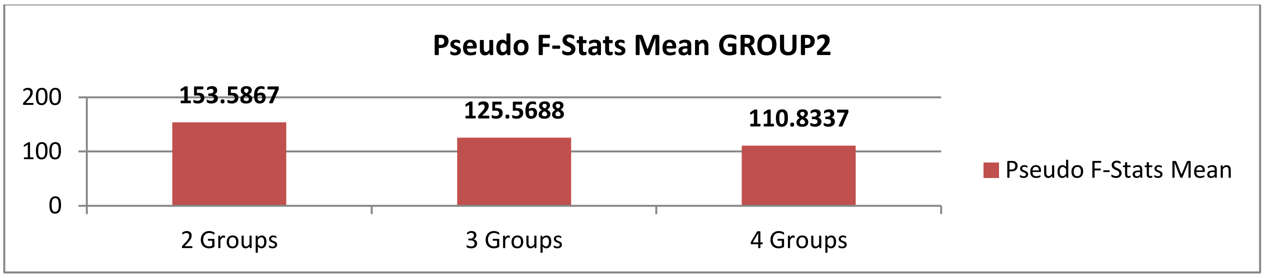


Figure 55. Pseudo F-Statistics Mean for partitioning the First Hierarchical Level Group 2 in sub-groups.

GROUP2	Entities belonging to each group			
K number	1 group	2 group	3 group	4 group
2	855	962	-	-
3	511	716	590	-
4	238	527	598	454

Table 22. Entities belonging to each sub-group of the First Hierarchical Level Group 2.

GROUP2	Assessment Parameters				
K number	Mean	Max value	Min value	Max Gap from Mean	Max intra-cluster Gap
2	908.5	962	855	|53.5|	107
3	605.67	716	511	|110.34|	205
4	454.25	598	238	|216.25|	360

Table 23. Results of assessment parameters used to evaluate sub-groups clustering results.

The evaluation of parameters indicates also in this case the K=2 solution as the most feasible for the second level partitioning.

The iterative partitioning of the first level Group 3 through the '*grouping analysis*' tool results in the following set of parameters, shown in Figure 56 and Table 24 and 25.

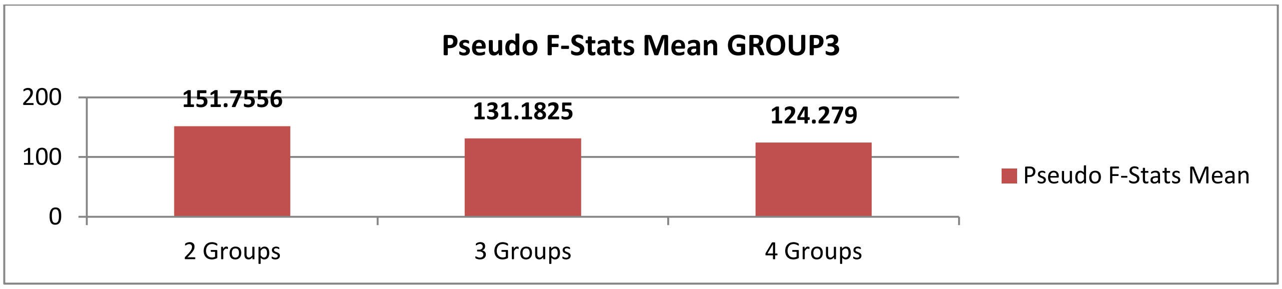


Figure 56. Pseudo F-Statistics Mean for partitioning the First Hierarchical Level Group 3 in sub-groups.

GROUP3	Entities belonging to each group			
K number	1 group	2 group	3 group	4 group
2	2920	6	-	-
3	1331	1584	11	-
4	1335	919	3	669

Table 24. Entities belonging to each sub-group of the First Hierarchical Level Group 3.

GROUP3	Assessment Parameters				
K number	Mean	Max value	Min value	Max Gap from Mean	Max intra-cluster Gap
2	1463	2920	6	|1457|	2914
3	975.33	1584	11	|608.67|	1573
4	731.5	1335	3	|603.5|	1332

Table 25. Results of assessment parameters used to evaluate sub-groups clustering results.

The evaluation of resulting parameters for partitioning the Group 3 suggests to use the solution K=4, which presents the most satisfying cluster size parameters.

Along the same vein, the results of Group 4 partitioning lead toward the identification of 4 sub-groups (K=4) as demonstrated by the following results. As a matter of fact, the highest Pseudo F-Stats value (Fig.57) indicates a suitable and compact clustering solution, meanwhile the cluster size assessment parameters show a similar partitioning result (Table 26 and Table 27).

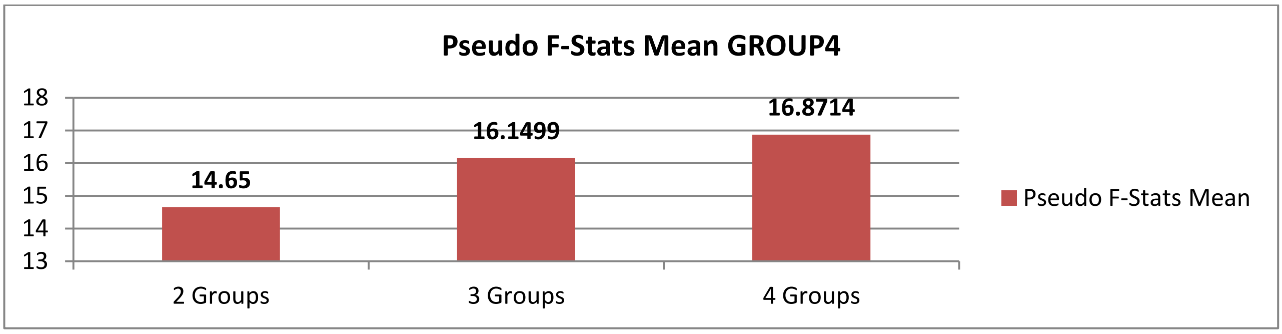


Figure 57. Pseudo F-Statistics Mean for partitioning the First Hierarchical Level Group 4 in sub-groups.

GROUP4	Entities belonging to each group			
K number	**1 group**	**2 group**	**3 group**	**4 group**
2	246	10	-	-
3	64	191	1	-
4	8	9	209	30

Table 26. Entities belonging to each sub-group of the First Hierarchical Level Group 4.

GROUP4	Assessment Parameters				
K number	**Mean**	**Max value**	**Min value**	**Max Gap from Mean**	**Max intra-cluster Gap**
2	128	246	10	|118|	236
3	85.34	191	1	|105.67|	190
4	64	209	8	|145|	201

Table 27. Results of assessment parameters used to evaluate sub-groups clustering results.

The first hierarchical level Group5 partitioning is evaluated with the same procedure and leads to the following results in terms of Pseudo F-Statistics (Fig.58) and clusters size assessment (Table 28 and 29).

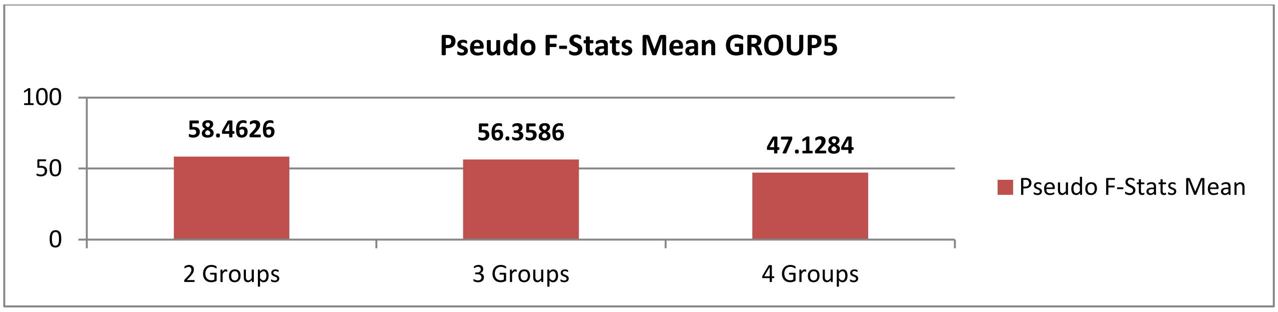


Figure 58. Pseudo F-Statistics Mean for partitioning the First Hierarchical Level Group 5 in sub-groups.

GROUP5	Entities belonging to each group			
K number	**1 group**	**2 group**	**3 group**	**4 group**
2	169	733	-	-
3	429	114	359	-
4	140	362	393	7

Table 28. Entities belonging to each sub-group of the First Hierarchical Level Group 5.

GROUP5	Assessment Parameters				
K number	Mean	Max value	Min value	Max Gap from Mean	Max intra-cluster Gap
2	451	733	169	|282|	564
3	300.67	429	114	|186.67|	315
4	225.5	393	7	|218.5|	386

Table 29. Results of assessment parameters used to evaluate sub-groups clustering results.

Finally, the iterative partitioning of the Group 6 identifies K=3 as the most feasible clustering solution in terms of Pseudo F-Statistics results, as provided in Figure 59. Notwithstanding the clusters size assessment parameters (Table 30 and Table 31) should suggest the use of the K=2 solution, in this case the inherent high intra-cluster similarities and inter-cluster differences foster the selection of 3 sub-groups.

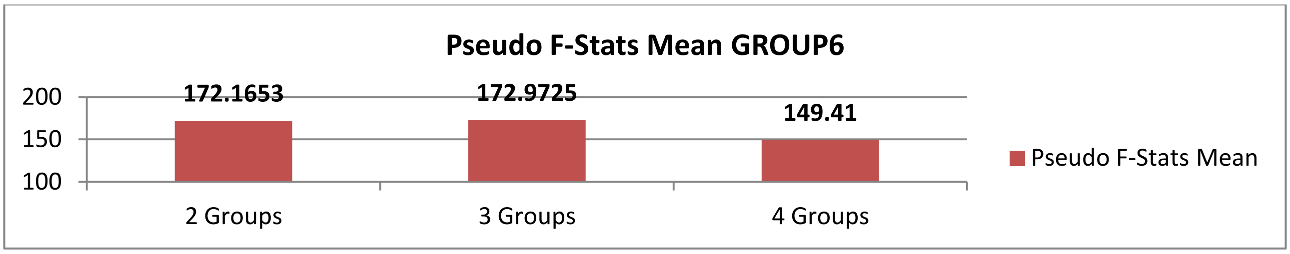


Figure 59. Pseudo F-Statistics Mean for partitioning the First Hierarchical Level Group 6 in sub-groups.

GROUP6	Entities belonging to each group			
K number	1 group	2 group	3 group	4 group
2	1278	1755	-	-
3	1138	313	1582	-
4	1575	305	1144	9

Table 30. Entities belonging to each sub-group of the First Hierarchical Level Group 6.

GROUP6	Assessment Parameters				
K number	Mean	Max value	Min value	Max Gap from Mean	Max intra-cluster Gap
2	1516.5	1755	1278	|238.5|	477
3	1011	1582	1138	|698|	1269
4	758.25	1575	9	|749.25|	1566

Table 31. Results of assessment parameters used to evaluate sub-groups clustering results.

The creation of the second hierarchical level concludes the partition of the Sardinian census tracts in order to develop the geodemographic classification. In summary, the first hierarchical level is composed by 6 Groups, meanwhile the second level consists of 18 sub-groups, which further stress the main features of resident people relying upon the 43 socio-economic identified variables. Table 32 summarizes the structure of the Sardinia geodemographic classification, while Figure 60 shows the obtained results.

SARDINIA GEODEMOGRAPHIC CLASSIFICATION	
FIRST HIERARCHICAL LEVEL	**SECOND HIERARCHICAL LEVEL**
Group 1	Sub-Group 1; Sub-Group 2
Group 2	Sub-Group 3; Sub-Group 4
Group 3	Sub-Group 5; Sub-Group 6; Sub-Group 7; Sub-Group 8
Group 4	Sub-Group 9; Sub-Group 10; Sub-Group 11; Sub-Group 12
Group 5	Sub-Group 13; Sub-Group 14; Sub-Group 15
Group 6	Sub-Group 16; Sub-Group 17; Sub-Group 18

Table 32. Geodemographic Classification structure.

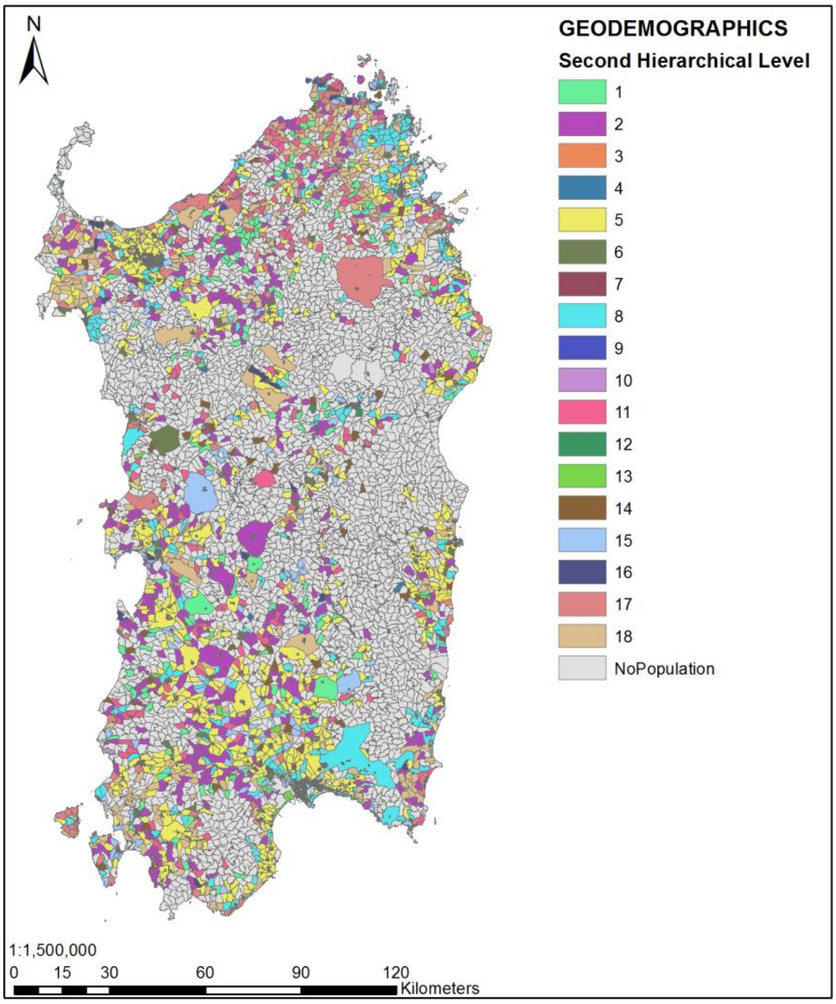


Figure 60. Geodemographic Classification: second hierarchical level.

7.5.4 Clusters profiling

The last step of the geodemographic classification is the clusters profiling, that is the assignment of a significant natural language label to each identified group and sub-group. The labels are assigned on the basis of the inherent features, which mainly enable to distinguish a cluster from its fellows. Commonly, the clusters profiling is used to foster an immediate comprehension of the groups characteristics; however the labels are not easy to produce, inasmuch they must not contradict official information and have to avoid the use of already established names (Vickers and Rees, 2007). The main argument for labeling a cluster is to easier the geodemographic classification use and the understanding of results.

In order to find suitable labels for the identified clusters at the first and second hierarchical level, the methodological approach relies upon the use of radial plots and histograms, which enable the direct observation and comparison of clusters characteristics. The first hierarchical level groups are processed by comparing the intra-group variables' means of the group census tracks with the variables' means of all the census tracks at the regional level, identifying the most distinguishing features of each group. Similarly, the second hierarchical level sub-groups are investigated by comparing the intra-subgroup variables' means of the sub-group census tracks with the variables' means of all their father groups census tracks. The analysis is based upon the examination of the variables values for each group, identifying which variables expose high and low values, consecutively selecting the most important ones in terms of mean' deviation to define a label.

The clusters profiling results are shown in Table 33, meanwhile the explanatory sheets for each group and sub-group are provided in attachment to the thesis, identifying quantitatively the most distinguishing features.

CLUSTERS PROFILING RESULTS	
First hierarchical level Groups	**Second hierarchical level sub-groups**
1. Single farmers in independent buildings	[1] 1.1 Elder single farmers [2] 1.2 Young single farmers
2. Residents of condominium and multi-apartment buildings	[3] 2.1 Elder blue-collars [4] 2.2 Middle-class families
3. Recent buildings residents	[5] 3.1 Young Blue-collars families [6] 3.2 Young prosperous families [7] 3.3 Young entrepreneurs families [8] 3.4 Tradesman families
4. Retired in small buildings	[9] 4.1 Retired businessmen [10] 4.2 Low-education retired on-a-budget [11] 4.3 Well-off retired families [12] 4.4 On-a-budget couples
5. Tenants	[13] 5.1 Middle-class singles [14] 5.2 Single white-collars [15] 5.3 Young well-off families
6. Old buildings residents	[16] 6.1 Elder working couples [17] 6.2 Tradesman retired families [18] 6.3 Blue-collars retired families

Table 33. Clusters profiling results.

The identified labels are assigned on the base of most distinguishing features arising from the investigation of groups and sub-groups variables. However, naming the clusters is a difficult process that may be limited by many hurdles in term of descriptive power and names' significance. Indeed, it is important to be aware that the proposed descriptive labels are not able to capture more than a small part of the real clusters characteristics and should be used with great caution in order to avoid the loss of information. In addition, the assigned labels are constantly updated in order to better fit with the emerging characteristics of the analyzed census tracts. Hence, the proposed labels should be considered exclusively as a tentative approach to name the groups resulting from the geodemographic classification. Despite the number of potential issues and limitations arising from the clusters' natural language description, the labels are used in the next methodological step as basis for enabling the SMGI user profiling.

7.6 User Profiling

After the conclusion of Sardinian census tracts geodemographic classification, conducted both at the first and the second hierarchical level, the methodological approach deals with the user profiling step. The user profiling is carried out by assigning each user to one of the 18 created categories, taking advantage of the previously identified residential clusters. Whereas the geodemographic groups are provided for the Sardinia territory exclusively, the users residing outside Sardinia are profiled as 'Italian tourists', if they are resident in Italy, or as 'foreigner tourists' if their residential locations are located outside the Italian boundaries.

This approach is developed in order to investigate potential underlying spatial and temporal patterns or dynamics due to the preferences of specific population groups within the study area. The methodology aims are the development of focused analyses in order to investigate:

1) which are the favorite temporal patterns to frequent a specific place by a specific group;

2) which are the behavioral commonalities between different groups;

3) which places attract the major attention of specific groups; 4) when the groups' attention is focused on.

The figure 61 shows the user profiling results developed in GIS environment.

The profile assigned to each census tract is transferred to the users residing into that area thanks to the development of a spatial overlay analysis with the previously identified residential clusters. As a matter of fact, each residential cluster belongs to a specific user and the centroid of the cluster itself is part of a single census tract. Through a simple overlay analysis and a spatial join between the clusters centroids and the census tracts each dwelling user may be classified accordingly to the geodemographic sub-group of his residential area.

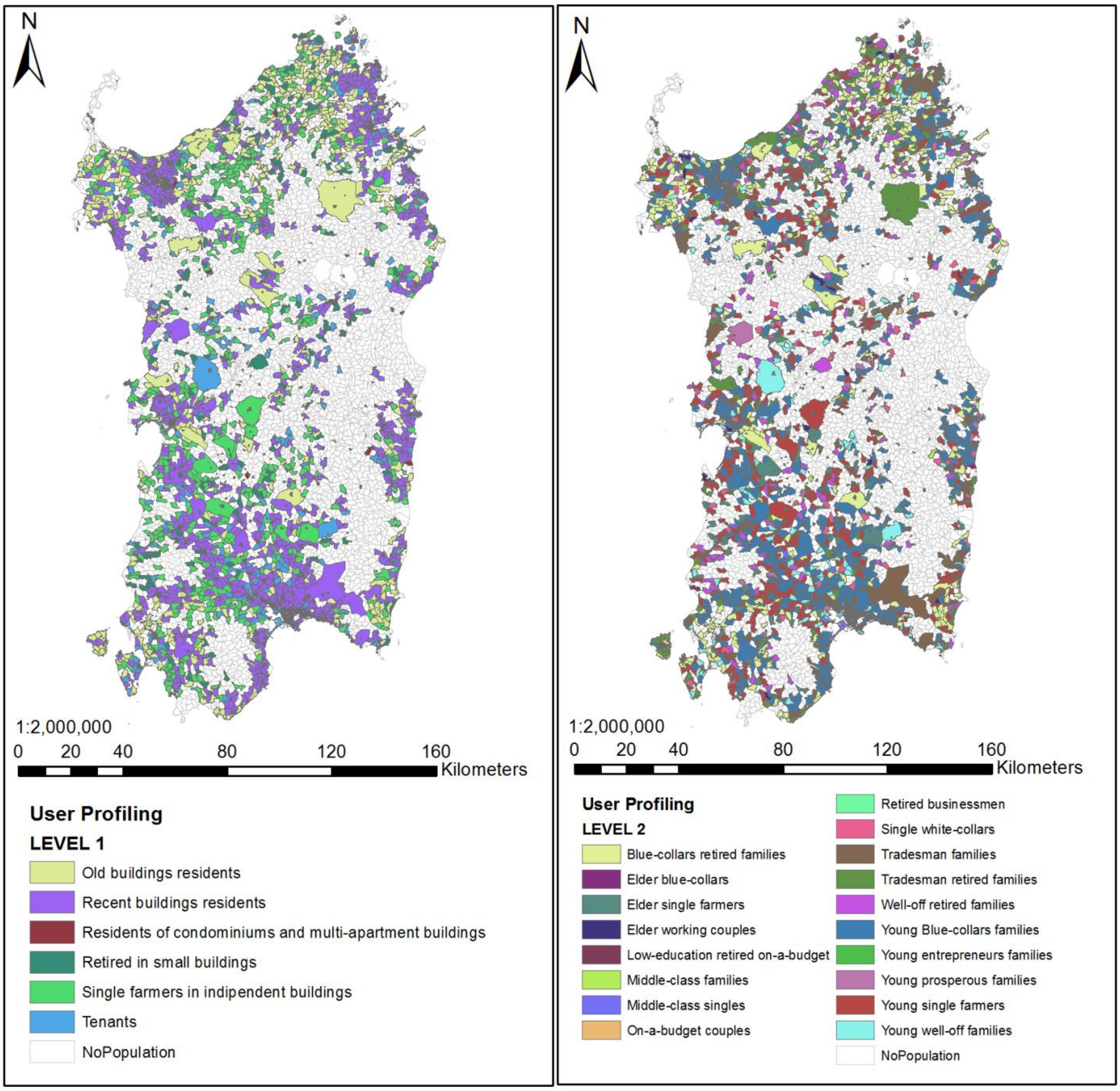


Figure 61. User profiling results.

The results of the overlay analysis, leading to the profiling of users contributing to the SMGI dataset, are provided in Table 34 and Table 35. The evaluation of residential clusters and profiled census tracts by means of spatial analyses results in the following user profiling, where each user is classified for the revealed residential provenance and geodemographic profile.

USER PROVENANCE	NUMBER OF USERS
Sardinia	1974
Italy	381
International	200

Table 34. Classification of users for provenance.

ID	GEODEMOGRAPHIC PROFILE	NUMBER OF USERS
[4]	Middle-class families	465
[6]	Young prosperous families	381
[16]	Elder working couples	265
[3]	Elder blue-collars	257
[18]	Blue-collars retired families	226
[5]	Young Blue-collars families	214
[8]	Tradesman families	85
[13]	Middle-class singles	47
[15]	Young well-off families	16
[17]	Tradesman retired families	7
[14]	Single white-collars	5
[7]	Young entrepreneurs families	2
[2]	Young single farmers	2
[12]	On-a-budget couples	1
[1]	Elder single farmers	1
[10]	Low-education retired on-a-budget	0
[9]	Retired businessmen	0
[11]	Well-off retired families	0
[I]	Italian tourists	381
[F]	Foreign tourists	200

Table 35. User profiling: users belonging to each identified category.

At the current stage, the assigned labels are still under revision in order to obtain a feasible descriptive characterization based upon the socio-demographic variables. In addition, a number of the 18 sub-groups present specific characteristics that may suggest the creation of a third level hierarchy that will be considered in future research. However, at present the research is conducted relying upon the aforementioned user profiling.

7.7 Multi-dimensional analyses on public space

The last methodological step concerns the development of multi-dimensional analyses on the 'Poetto beach and the Molentargius Park' study area exploiting the information obtained from the previous analyses. The analyses are conducted in order to investigate the users' behavior in space and time with the aim of characterizing and describing the public spaces according to the identified dynamics.

The proposed methodology allows the development of focused analyses on the spatial and temporal dimension of the SMGI dataset integrating both the information hailing from the user profiling and the availability of official information. In addition, in order to properly describe the particular features of the study area, a number of complementary SMGI extractions are conducted. The increased availability of information, made available by the described SMGI Analytics approach, eases the development of analyses which embrace all the dimensions of SMGI, namely space, time and user. A list of potential analytical methods may include as follow:

- public space analysis by user profiling:
  - foreign tourists dynamics analysis;
  - Italian tourists dynamics analysis;
  - resident users dynamics analysis;
- public space analysis by temporal dimension:
  - temporal analyses by user profiling;
- detection of Points of Interest within the study area:
  - identification of high interest POIs within general clusters;
  - identification of high interest POIs by user profiling;
  - comparison between places by user profiling or temporal period;
- user behavioral investigation:
  - analysis of similarities among different user profiles;
  - comparison of spatial patterns among different users profiles;
  - comparison of similarities and dissimilarities in temporal patterns of user groups;
- multi-dimensional analyses
  - a number of analyses coupling several of the aforementioned.

In the light of the above analytical opportunities, the research discusses the investigation and the results obtained from several analyses conducted on the study area in order to describe and characterize the user behaviors and the dynamics affecting the Poetto beach and the Molentargius Park areas.

7.7.1 Detection of Points of Interest within the public space

The analysis aims to identify the most appreciated and visited places and locals by the Instagram users within the study area. The identification of POIs in the area is carried out by integrating the previously identified clusters and then performing a complementary SMGI datasets extraction. As a matter of fact, the clusters emerging from the analysis conducted by means of the DB-SCAN algorithm enable to detect the areas attracting the major interest of users, but in order to gain further insights about the reasons behind the users preferences, it is necessary to perform further analyses at the local scale. Therefore, the clusters are enlarged from their centroid through the ArcGIS 'buffer tool' selecting a 25 meters radius and then investigated by coupling information obtained through SMGI datasets extracted from the Instagram Places and the Foursquare social media platforms, which may be considered as LBSNs that provide georeferenced information about POIs, as well as information about the number of user presences. A comparison between the two extracted complementary SMGI datasets is performed in order to assess and confirm the POI existence, while the attractiveness of each confirmed POI is evaluated though the metadata provided by Foursquare, relying upon the "number of check-in" and the "number of users that visited the place", so detecting the most attractive place within each cluster. Indeed, a specific POI is assigned to each cluster in terms of attractiveness, if it is embedded within the buffered cluster area, or by means of both attractiveness and spatial proximity, if a cluster does not contain any POI.

Moreover, the analysis concerns the investigation of places typology in order to provide the characterization of the study area relying on the most visited POIs. Afterwards, the POIs are examined in more detail at the local scale, identifying their names and evaluating how many clusters belong to a specific POI. As a matter of fact, several overlapping clusters may belong to the same POI.

The complementary SMGI extraction from Instagram Places and Foursquare is performed by means of SPATEXT tools, which allow the collection of POIs by specifying a spatial query. The extractions result in two dataset consisting of 771 POIs for Instagram and 177 POIs for Foursquare. Despite the notable difference in data volume, a comparison between the two datasets expose that the Instagram Places SMGI dataset may contain the same POI for multiple times due to the opportunity of users to label the places where they took photos. This capability leads toward the creation of the same POIs with different names. On the other hand, the dataset extracted from Foursquare is automatically and periodically checked and assessed by the social media platform itself, which deletes duplicates and assess the quality of contained POIs. In addition, the Foursquare dataset provides useful information regarding the typology of POIs and the number of users who visited and made a check-in in the considered locals, allowing the identification of most attractive POIs and enabling the study area characterization. The results of the complementary SMGI extraction are provided in Figure 62, exposing the spatial distribution of Instagram Places and Foursquare SMGI datasets.

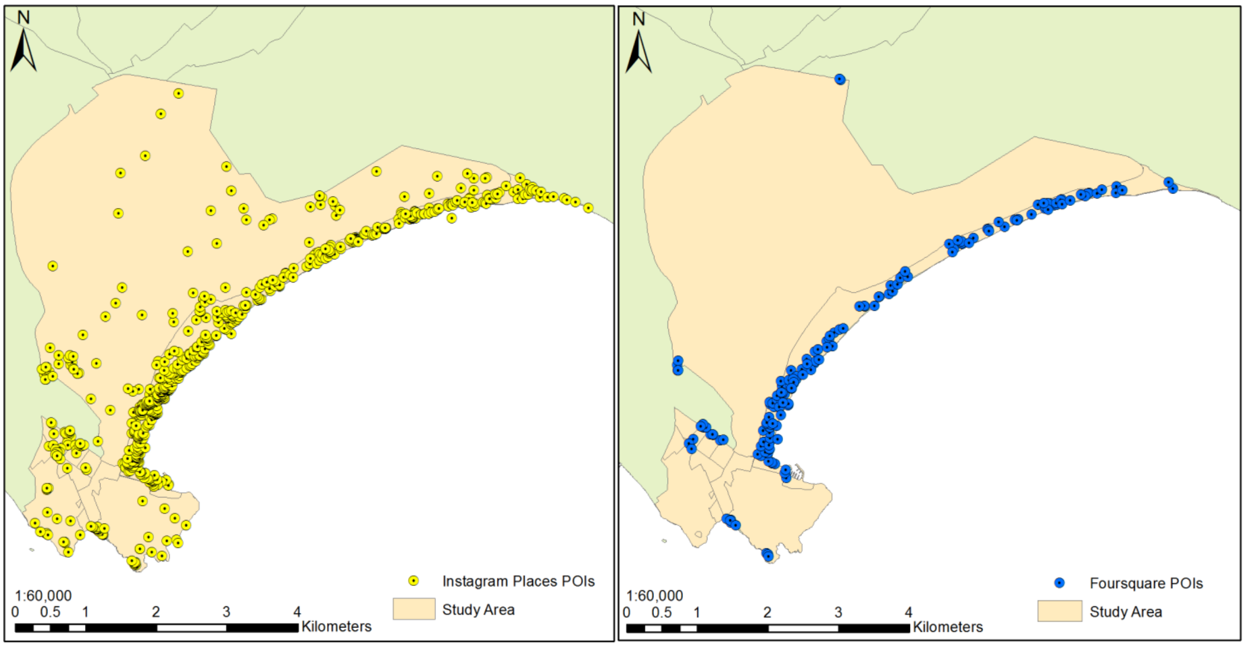


Figure 62. POIs obtained from the Instagram Places SMGI dataset and the Foursquare SMGI dataset.

The spatial analysis results are shown in Figure 63, highlighting the complementary SMGI datasets, the buffered clusters and the SMGI dataset of users' contribution who led to the clusters definition.

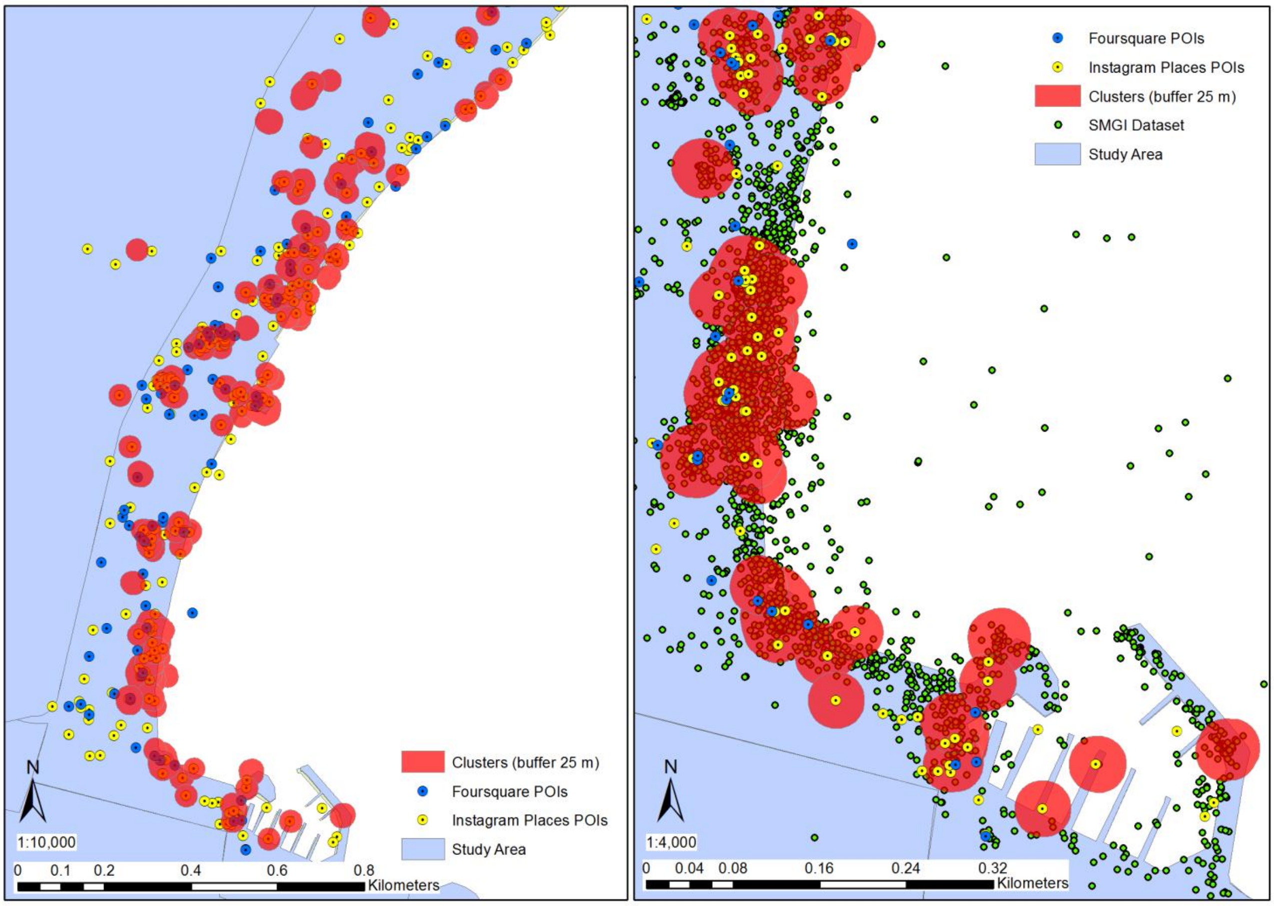


Figure 63. POIs identification: clusters, Instagram Places and Foursquare POIs and SMGI dataset.

The analysis of the most visited POIs for typology exploits the predefined categories provided by the Foursquare social network. Indeed, this LBSN enables users to georeference their position every time they visit (check-in) a specific place or locations and then automatically assigns each locations to a general and a specific category. Therefore, the different typologies of POIs within the study area, the most visited typologies of POIs according to users preferences, as obtained from clustering, and finally their identification at the local scale, are provided in Table 36, 37, and 38, respectively.

POIs Typology	**POIs number**	**POIs Typology**	**POIs number**
Beach	28	Café	15
Italian Restaurant	9	Restaurant	8
Bed & Breakfast	7	Bar	6
Other Nightlife	6	Nightclub	5
Surf Spot	5	Hotel	4
Cocktail Bar	3	Fast Food Restaurant	3
Food	3	Hospital	3
Pizza Place	3	Playground	3
Sandwich Place	3	Snack Place	3

Table 36. Main POIs categories within the study area as obtained from the Foursquare SMGI dataset.

POIs Typology	POIs number	POIs Typology	POIs number
Beach	38	Café	35
Italian Restaurant	16	Hospital	11
Restaurant	8	Bar	7
Cocktail Bar	7	Pizza Place	7
Snack Place	7	Playground	5
Spa	5	Bed and Breakfast	4
Cafeteria	4	Food court	4
Ice Cream Shop	4	Surf Spot	4
Dog Run	3	Fast Food Restaurant	3

Table 37. Most visited POIs typologies according to users' preferences within the study area.

POIs Name		POIs Typology	POIs Name	qnt.	POIs Typology
Calamosca	6	Beach	Il Lido	6	Beach
La Sella del Diavolo	6	Cafè	Ospedale Marino	6	Hospital
Centro Donna Binaghi	5	Hospital	La Pirata	5	Italian restaurant
Le Terrazze Calamosca	5	Cocktail Bar	Bobocono Beach	4	Ice Cream Shop
Calafighera	4	Beach	Capolinea	4	Beach
Emerson Cafè	4	Cafè	Kaìros	4	Cafè
La Lanterna Rossa	4	Cafè	La Marinella	4	Italian restaurant
La Rotondina	4	Cafeteria	Sa Sesta @ Poetto	4	Beach
Spinnaker	4	Food Court	Twist	4	Cafè
Il fico d'india	3	Cafè	La Paillote	3	Cocktail Bar

Table 38. Preferred POIs by users visits within the study area.

The obtained results demonstrate the potentialities of SMGI to elicit information related to the geography of places, fostering the POIs identification within the study area, enabling the characterization of the public place. As a matter of fact, the results show that the area is mainly visited for the presence of both the beach and a number of leisure places, namely cafè, restaurant and bar. However, the results stress also the presence of the hospital 'Ospedale Marino', which is located within the study area.

7.7.2 Public space dynamics analysis through user profiling

The multi-dimensional analysis discussed in this section aims to investigate the dynamics of a specific users group within the study area, namely the tourists users. The approach exploits of the SMGI dataset spatial distribution and the user profiling results to identify the most appreciated areas by tourists and investigate at the local scale which POIs mainly attract the attention and the visits of this population segment. Relying upon the user profiling results, the contributions of users, classified as 'Italian tourist' and 'Foreign tourist', are investigated in terms of spatial distribution and density by means of heat maps, developed through the 'kernel density' tool of ESRI ArcGIS. The tool calculates a magnitude per unit area from the SMGI features using a kernel function for adjusting a smoothly tapered surface to each point in the SMGI dataset.

The approach is based upon the same procedure used to evaluate the existence of potential spatial patterns of interest according to different periods of time during the explorative spatial-temporal step. However, in this case, the final results visualization relies on both the Natural Breaks (Jenks) and the

Quantile visualization. On the one hand, the Natural Break (Jenks), or natural interval, relies on an algorithm able to detect the natural grouping of data to define the final classes. The resulting classes present the minimum intra-group variance and the maximum inter-groups variance. On the other hand, the Quantile visualization defines the resulting classes in order to obtain a number of classes exposing the same number of contained features. The choice to present the results through two different visualization methods is made in order to ease the visual assessment of areas which expose the major attractiveness for tourists within the study area. The 'kernel density' results on the SMGI contributions related to 'Italian tourists' and 'Foreign tourists' are shown in Figure 64, identifying the places within the study area attracting the major interest.

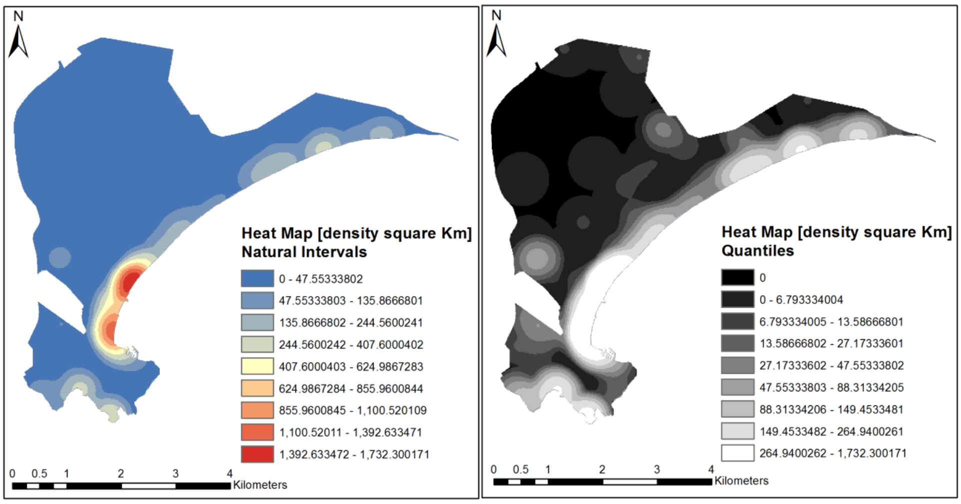


Figure 64. Attractiveness heat maps of the study area through 'Kernel density' tool.

The attractiveness heat maps, developed by means of kernel density analysis, highlight the major density of tourists' contributions within specific places that are further investigated considering both the study area's clusters, the SMGI contributions and the complementary Instagram Places and Foursquare SMGI datasets containing the POIs. The results are shown in Figure 65.

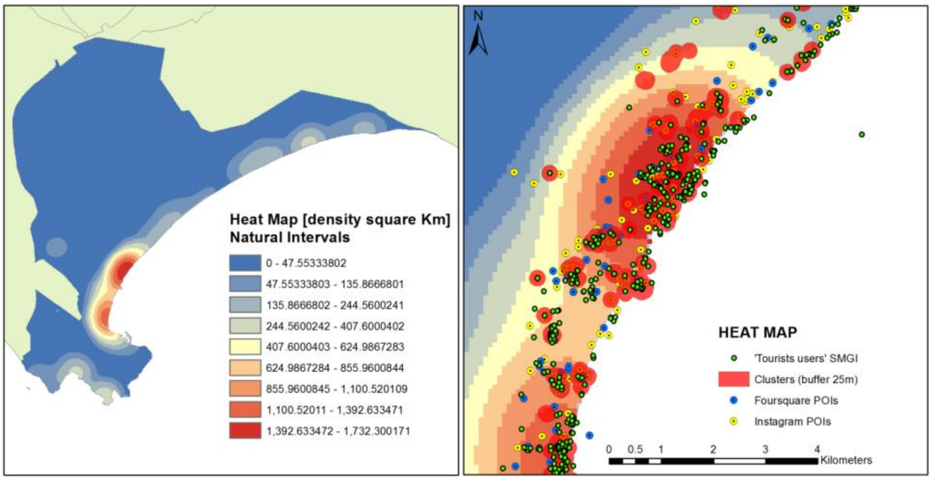


Figure 65. Identification of areas exposing the higher density of SMGI contributions by tourists.

The higher density of SMGI contributions is located in specific places within the study area and leads toward the investigation of the POIs comprised within these areas. In order to identify the POIs attracting the major interest of tourists, an overlay analysis is conducted between the heat map and the Foursquare SMGI dataset. The results of the spatial analysis identify 5 different POIs within the areas exposing the major density, which are listed in Table 39.

POIs name	POIs typology
Malibù	Cafè
Bobocono Beach	Ice Cream Shop
Oasi Cafè	Cafè
Emerson Café	Cafè
Palm Beach	Snack

Table 39. Top 5 favorite POIs by the tourists within the study area.

The final results demonstrate the capabilities of the SMGI Analytics methodology to inquiry specific areas at the local scale in order to gain useful insights about the users' behavior and preferences. This analysis exposes the tourists' trend to visit a number of services along the seaside of the Poetto beach.

7.7.3 User profiling to compare places

The third multi-dimensional analysis concerns the investigation of two bordering places within the study area in order to analyze potential dissimilarities in the profile of users who visit these places. The investigated places are notably different in nature and sizes. As a matter of fact, the first one is the 'Lido', namely a seaside establishment that offers a number of different services to customers to spend the daily hours in the Poetto beach, as well as, a leisure night local / discotheque / bar that offers the opportunity to spend the night to its customers. On the other hand, the second investigated place is the bordering 'free beach', a little strip of beach without services that is usually lived by users to spend the day at the Poetto beach without the requirement to expend money to be visited.

For their inherent differences, the analysis aims at identifying differences both in terms of temporal patterns, users profiles and users provenance, relying upon the investigation of spatial, temporal and user dimension of Instagram SMGI. Immediately, an explorative analysis exposes the different degree of attractiveness offered by the two areas: the 'Lido' area comprises 2466 contributions, meanwhile the 'free beach' area presents exclusively 36 Instagram photos, denoting the importance played by customer services to the Instagram users for visiting and staying in a place. In addition, the small number of contributions in the 'free beach' may be explained due to the limited size of the area. Nevertheless, despite the volume difference in the two samples, the analysis is carried out to investigate further potential dissimilarities. The figure 66 shows the spatial limits of the two considered places, their contained contributions and the complementary Foursquare SMGI dataset.

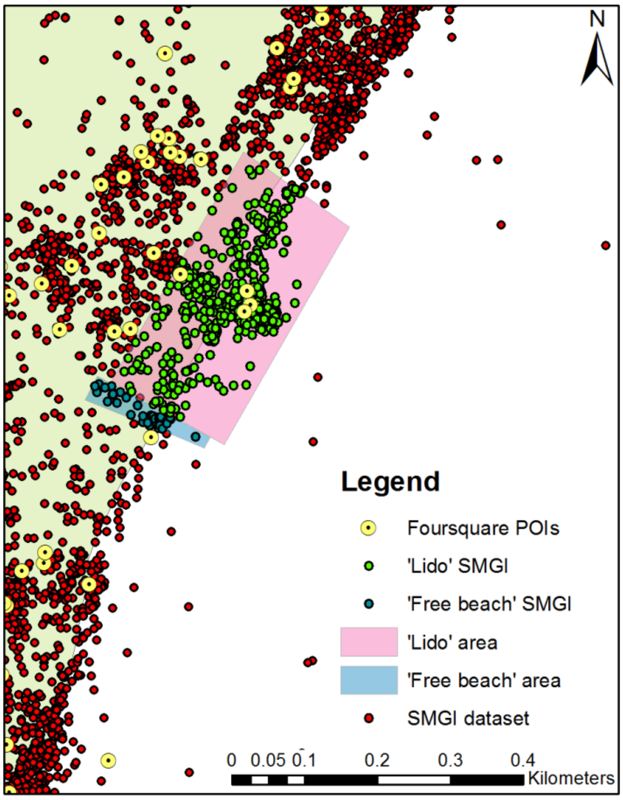


Figure 66. Identification of ‘Lido’ and ‘Free beach’ areas.

At this stage, the analysis deals with the investigation of differences in user profiles that visit the two different places exploiting the results of user profiling. As a matter of fact, any SMGI is produced by an user that was profiled in the previous step, disclosing the opportunity to investigate which users groups visit or prefer a certain area. The results show that the ‘Italian tourists’ or ‘Foreign tourists’ are absent in the ‘Free beach’ area, raising interesting questions about the advantages that may be offered by customer services for attracting the interest of this user group. The results of comparison between user profiles for the two areas are presented in Figure 67 and Table 40.

USER PROFILE	Users in ‘Free beach’ area	Users in ‘Lido’ area
Middle-class families	7	130
Young prosperous families	3	76
Elder working couples	2	51
Elder blue-collars	2	61
Young Blue-collars families	0	47
Blue-collars retired families	1	38
Middle-class singles	0	10
Tradesman families	0	19
Italian tourists	0	23
Foreign tourists	0	8
Young well-off families	0	5
Single white-collars	0	1
Young single farmers	0	1

Table 40. Comparison between the ‘Free beach’ and the ‘Lido’ area in terms of user profiles.

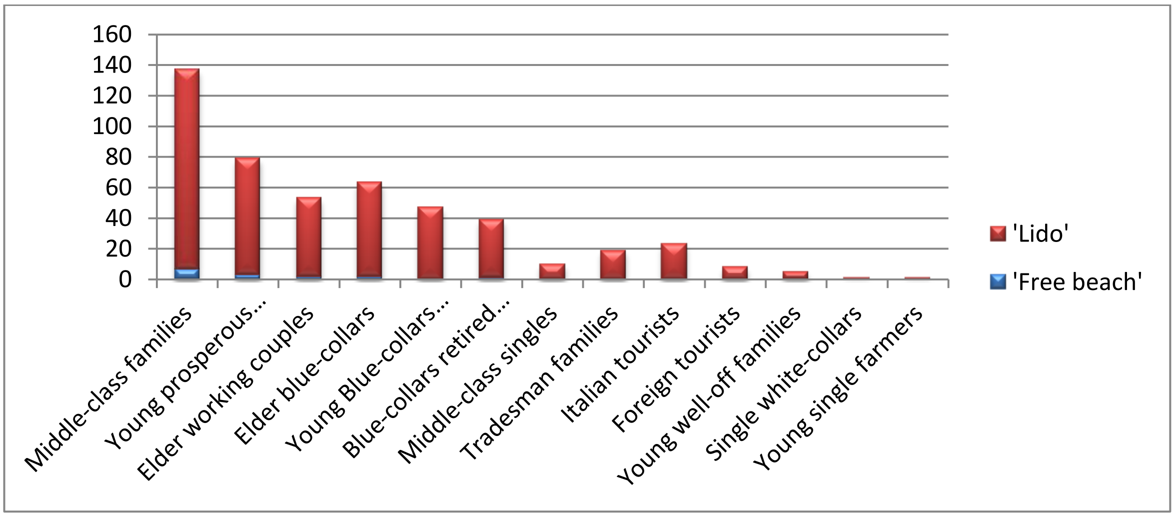


Figure 67. Differences in users profiles between the 'Free beach' and the 'Lido' area.

After the conclusion of the comparison between the areas in terms of user profiles, the analysis concerns the investigation of difference in temporal patterns. The temporal analysis is conducted considering the SMGI monthly distribution and the daily trends, while the results are provided in Figure 68 and Figure 69, respectively.

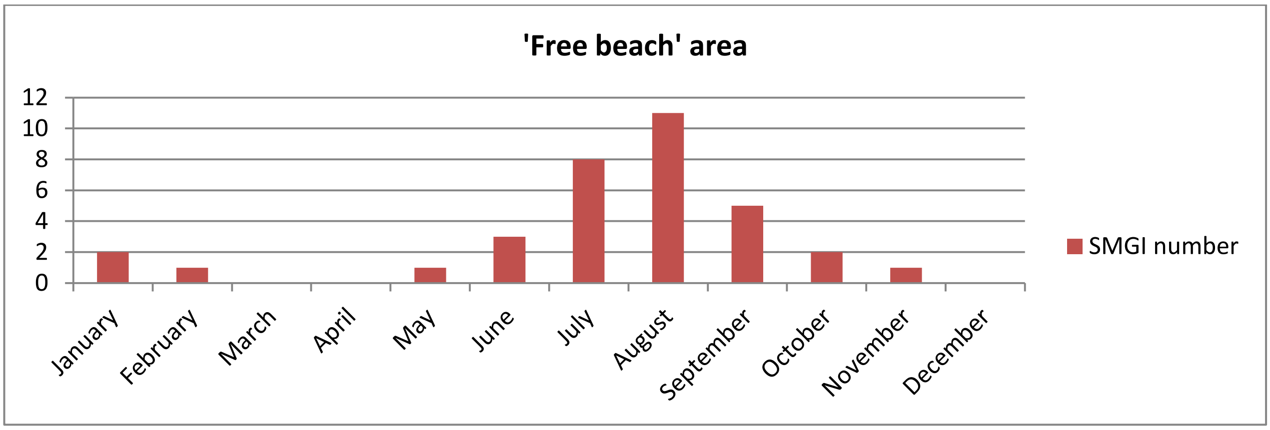


Figure 68. Monthly distribution of SMGI in the 'Free beach' area.

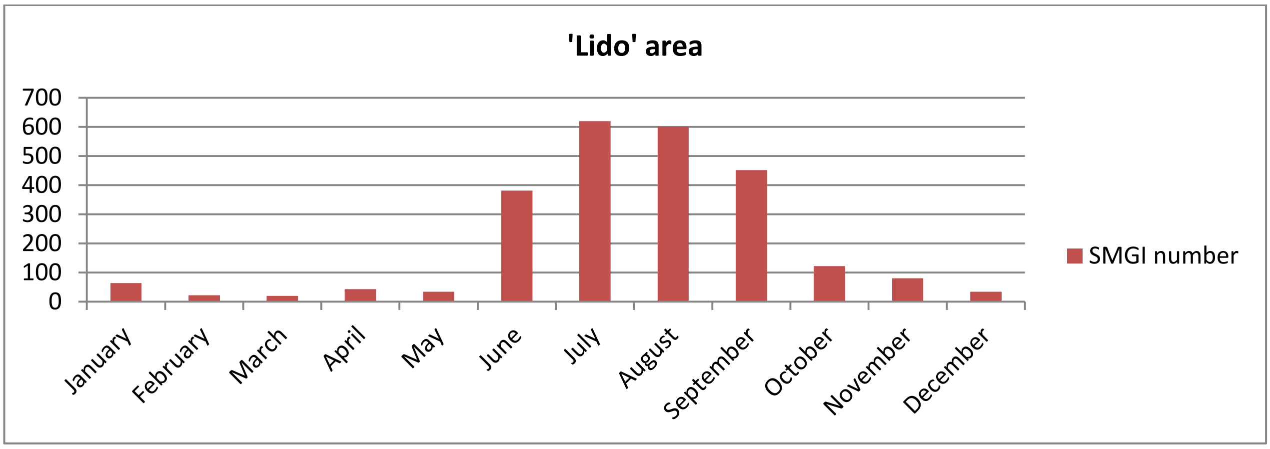


Figure 69. Monthly distribution of SMGI in the 'Lido' area.

The results of monthly temporal investigation expose the differences between the two considered areas. Notwithstanding the differences in SMGI volume, the analysis stresses the presence of users during all the year in the ‘Lido’ area, denoting the multi-purpose nature of the place that offers a set of differenced services to meet the customers’ expectation during all the year. On the contrary, the ‘Free beach’ area is mainly lived only during the summer and presents several periods without SMGI contributions.

The same observed phenomenon is visible analyzing the daily temporal trends for the two areas, as shown in Figure 70 and Figure 71. Indeed, in spite of normal fluctuations in the trend, SMGI contributions are shared by users during all the day in the ‘Lido’ area. In particular, the high number of SMGI uploaded by users during the night hours denote the different services provided by the place to its customers, which may live the place both during day for the beach but also during night for the offered leisure activities. Similarly to the monthly distribution, the ‘Free beach’ presents different periods without contributions, particularly during the night, highlight the lack of service in the area and the potential consequent absence of visitors.

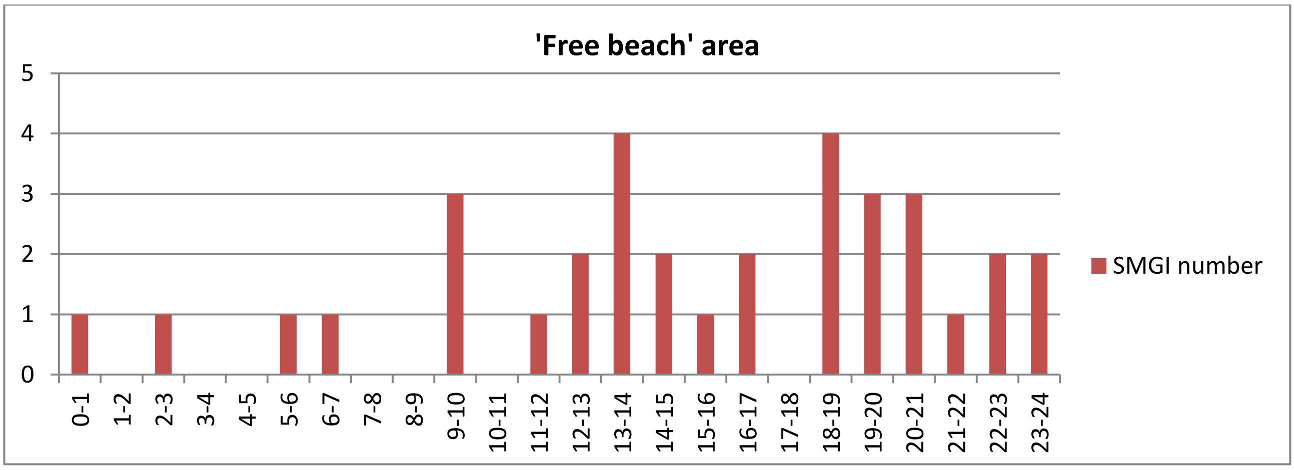


Figure 70. Daily trend of SMGI in the ‘Free beach’ area.

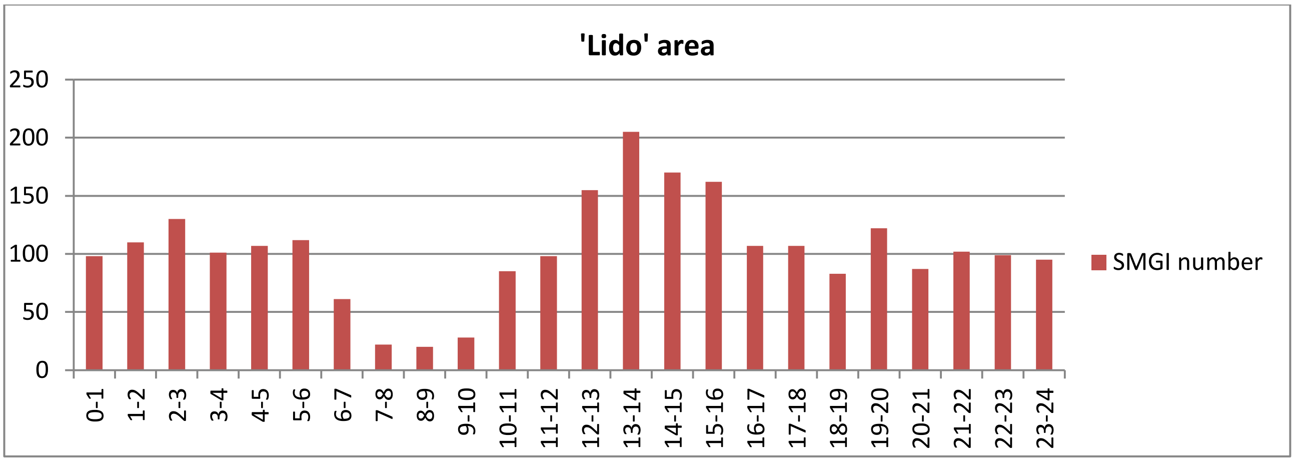


Figure 71. Daily trend of SMGI in the ‘Free beach’ area.

A focused analysis of the temporal trends denote the minimum number of contributions for the ‘Lido’ area during the periods 07:00-08:00 AM and 08:00-09:00 AM, that may be due to the functional change in services offered in the area. Indeed, during these periods, the night services are just closed and the day services are starting to open.

Finally, the last analysis carried out on the two areas concerns the investigation of users’ provenance. In fact, the SMGI Analytics allowed the detection of users residential areas, allowing the study of their provenance. The aim of this analysis is to study if the presence of different customer services in the ‘Lido’ area attracts visitors from distant areas more than the ‘Free beach’ area, which does not offer any kind of customer service. The results of the spatial analysis on the users’ provenance are shown in Figure 72, denoting notable differences between the two studied areas.

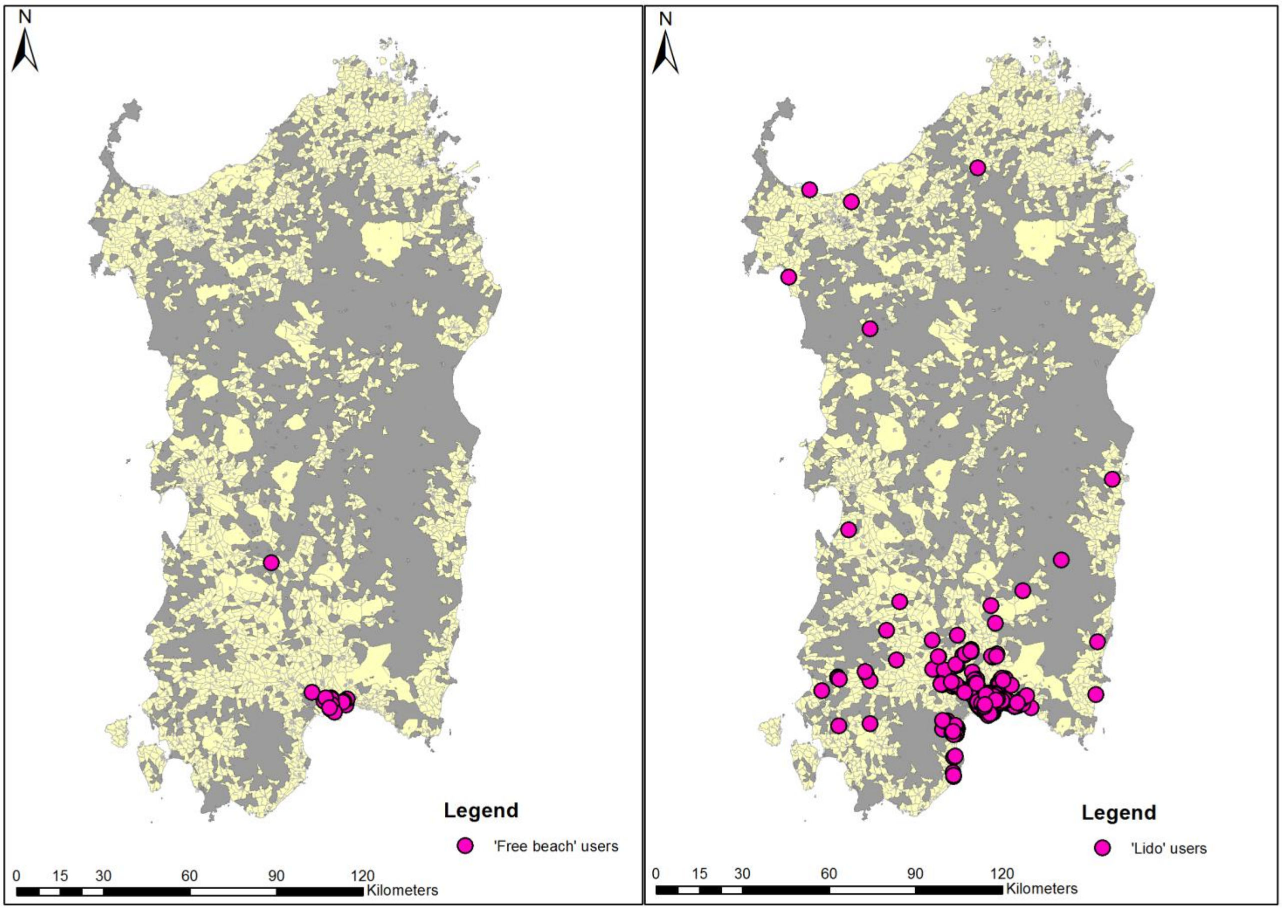


Figure 72. Users provenance for the ‘Free beach’ and the ‘Lido’ area.

The obtained results show that the ‘Free beach’ area is visited exclusively by 15 users in the SMGI dataset, that are mainly residents in the Cagliari municipality or in bordering municipalities to the investigated area. On the other hand, the ‘Lido’ area is visited by 470 users, and the 40% of them (188 users) are residents outside the Cagliari municipality or its bordering municipalities. This phenomenon confirms the notable attractiveness of the ‘Lido’ area, probably due to the specific characteristics of the place, which offers several combined services able to satisfy requirements of different user groups, who decide to move and to visit the place. In addition, the 'Lido' area capability to attract users is confirmed by the presence of 31 tourists users, namely the 6.60% of the sample, that shared Instagram contributions when visiting the place.

7.7.4 Temporal investigations to characterize places

The fourth multi-dimensional analysis concerns the characterization of a place in the study area through a set of temporal investigations aiming at identifying potential differences in daily use patterns. The analysis concerns the 'Lido' area, that is a seaside establishment offering several services to customers for spending their day or night hours in the Poetto beach. Similarly to the previous analysis, the 'Lido' area is notably attractive to users and about 2500 SMGI were uploaded in one year within its boundaries by over 400 users. However, the SMGI volume might be produced with different rates during certain time intervals, potentially denoting different customer groups' preferences in using the place only for particular purposes or during specific periods.

In the light of these considerations, the 'Lido' area is investigated by means of explorative temporal analysis aimed at identifying potential differences in spatial distributions and user profiles. First of all, the 'Lido' area is evaluated in terms of exploratory spatial analysis, easing the identification of both day/night patterns and density of contributions, so figuring areas attracting the major interest. The SMGI contributions are classified in two categories, namely 'Day' (from 7.00 AM to 18:59 PM) and 'Night' (from 19:00 PM to 6:59 AM), which show similar data volume percentages that are 50.36% and 49.64%, respectively. The results of exploratory analysis related to spatial distribution and contributions density are shown in figure 73 A and B.

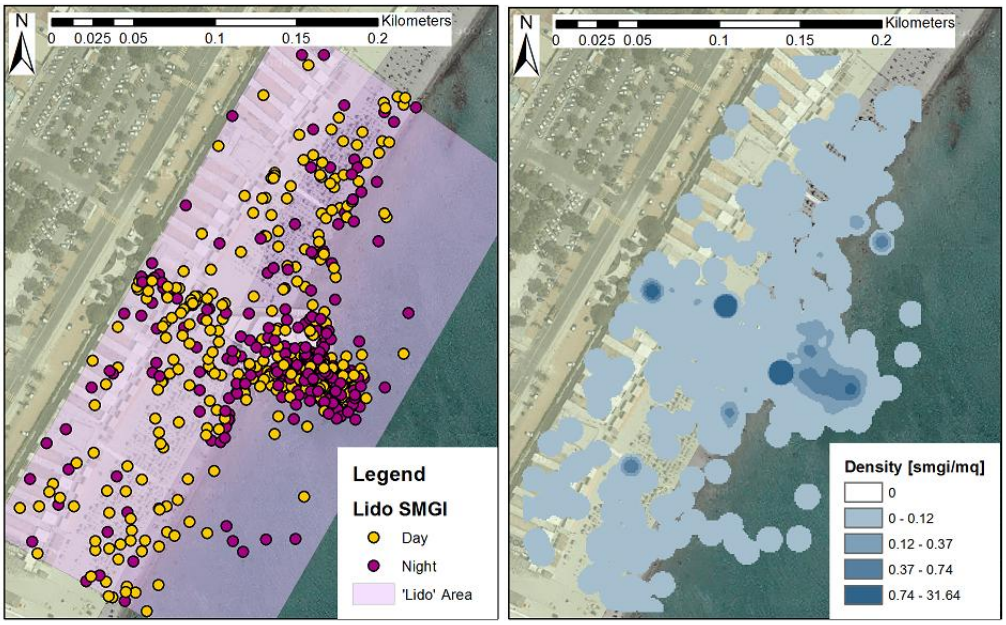


Figure 73. (A) 'Lido' Day and Night SMGI spatial distribution; (B) 'Lido' SMGI kernel density.

Following the first exploratory spatial investigation, the analysis concerns the study of the SMGI distribution in the area during different temporal periods. The temporal analysis is based upon three different temporal periods, namely the ‘day’ period, from 7:00 AM to 18:59 PM, the ‘night’ period, from 19:00 PM to 6:59 AM, and finally the ‘night life’ period, which is intended from 0:00 AM to 5:00 PM. The aim of the analysis is to identify potential differences in SMGI spatial distribution related to these different intervals, so eliciting differences in behaviors and preferences of 'Lido' users due to certain time periods. In particular, the third interval, namely ‘night life’, is selected with reference to the particular characteristics of the ‘Lido’ area itself, which offers a leisure services, such as discotheque and bar, in the night hours during the summer.

The analyses conducted on the three temporal periods highlight a notable difference in the use of the area. As a matter of fact, during the ‘day’ period the SMGI contributions are extensively distributed in the whole area, meanwhile, during the ‘night’ period and particularly in the ‘night life’ period, the contributions are increasingly grouped near the discotheque / bar service and near the entrance of the local, represented by the blue spots on the maps (figure 74 C). The results of the analysis are show in figure 74 A, B and C.

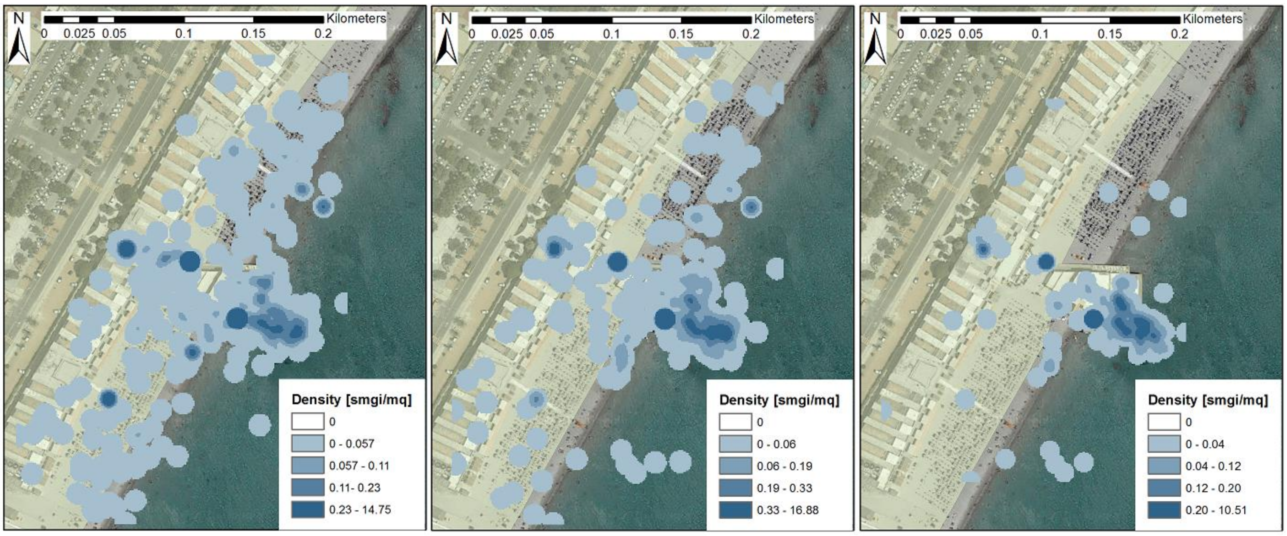


Figure 74. SMGI contributions during (A) ‘day’ period, (B) ‘night’ period, (C) ‘night life’ period.

Then, the analysis takes advantage of the temporal dimension of SMGI in order to investigate what period of the year is the most attractive to frequent the ‘Lido’ area during the day and the night. Not surprisingly, the temporal analysis identifies the summer season (from June to September) as the period when the ‘Lido’ exposes the highest attractiveness to users. In addition, the area is popular particularly during the weekend (Saturday and Sunday), as exposed from the temporal patterns of both the ‘day’ and ‘night’ periods. The seasonal phenomenon is probably caused by the geographic location of the place, which is located in the seaside, fostering the use during the summer period and limiting the service during the winter. On the other hand, both the increased leisure time of users and the growing activity of the discotheque / bar services during the weekends are probably the reasons of the identified daily patterns. The results of the temporal investigation concerning the most attractive period of the year are provided in Figure 75 A and B, while the weekly patterns are shown in Figure 76 for both day and night periods.

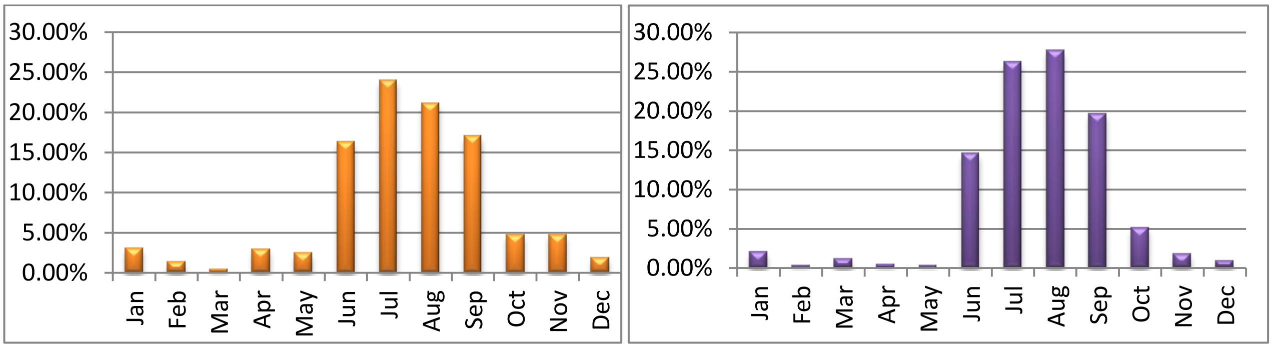


Figure 75. Rates of SMGI contributions for months during (A) 'day' period and (B) 'night' period.

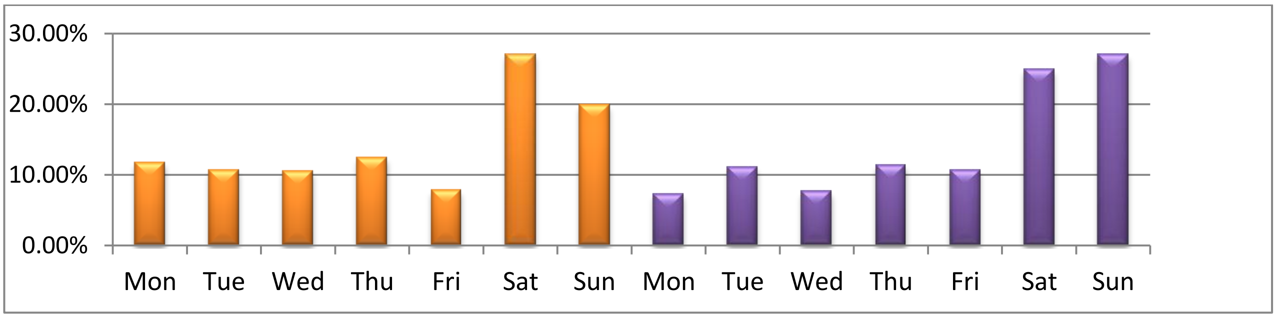


Figure 76. Rates of SMGI contributions for days of the week during 'day' period and 'night' period.

Finally, the last analysis conducted on the 'Lido' area concerns the investigation of users' provenance. This last analysis aims to investigate if the users frequenting the 'Lido' area during the 'day' period differ from the users of the 'night' period in terms of profiles or dwelling area. Figure 77 and Figure 78 report the user profiles identified for 'day' and 'night' periods, respectively.

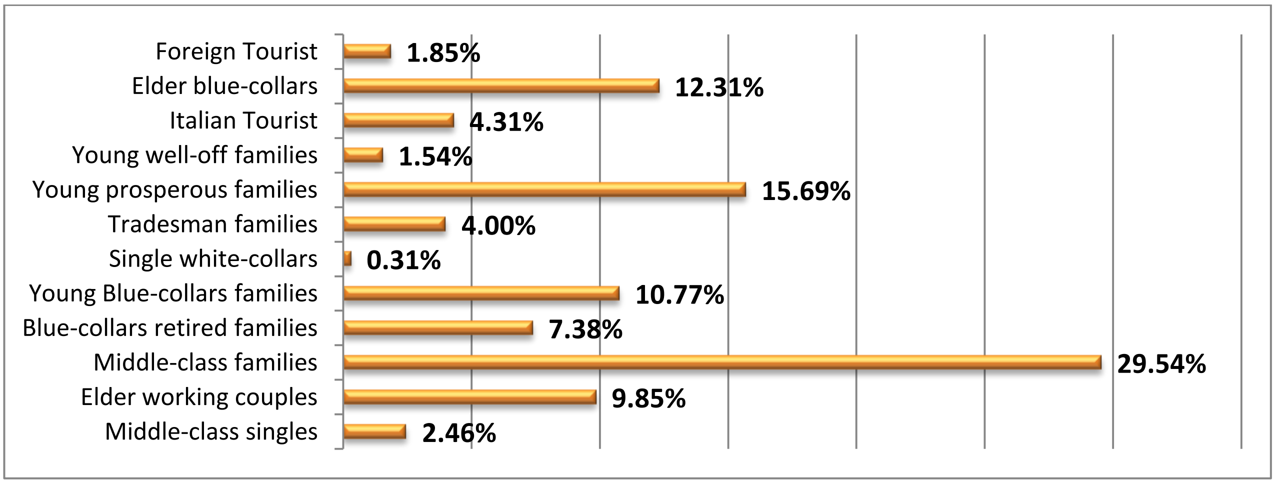


Figure 77. Rates of SMGI contributions for user profile during 'day' period.

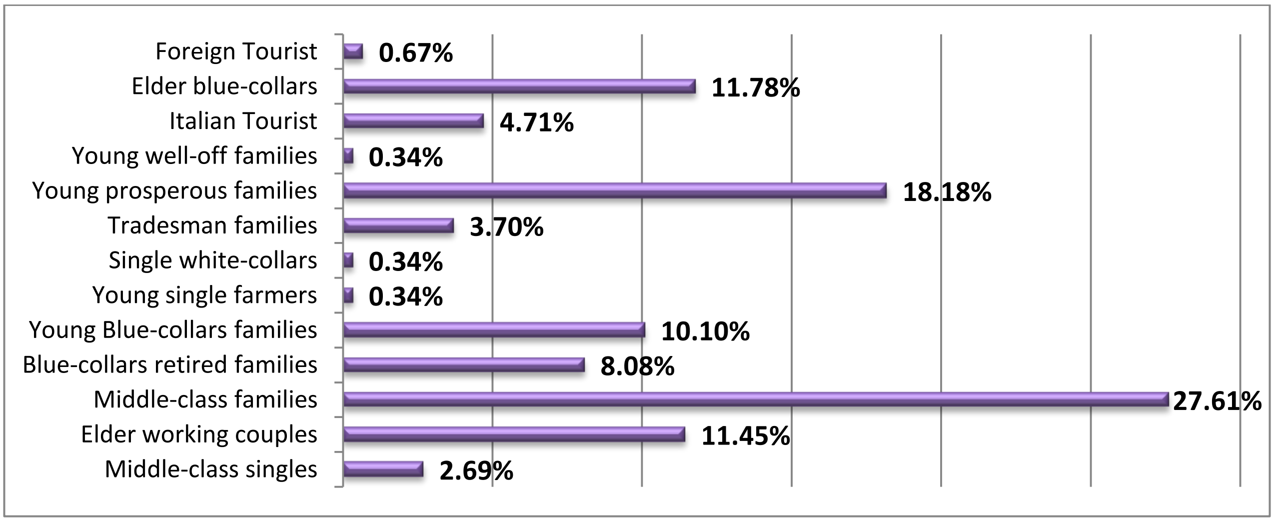


Figure 78. Rates of SMGI contributions for user profile during ‘night’ period.

The analysis carried on the user profiles shows a similar pattern for the two periods, demonstrating that the ‘Lido’ area is popular with the same population groups in spite of diverse temporal periods. Moreover, 152 users of the 470 analyzed are common for both the periods. The most notable difference is the presence of one more user profile, namely ‘young single farmers’, during the ‘night’ period.

Similarly, the spatial analysis of the users provenance shows similar distributions, as shown in figure 79 A and B, demonstrating the capability of the ‘Lido’ to attract users that inhabit outside the Cagliari municipality or its bordering municipalities, independently from the temporal period.

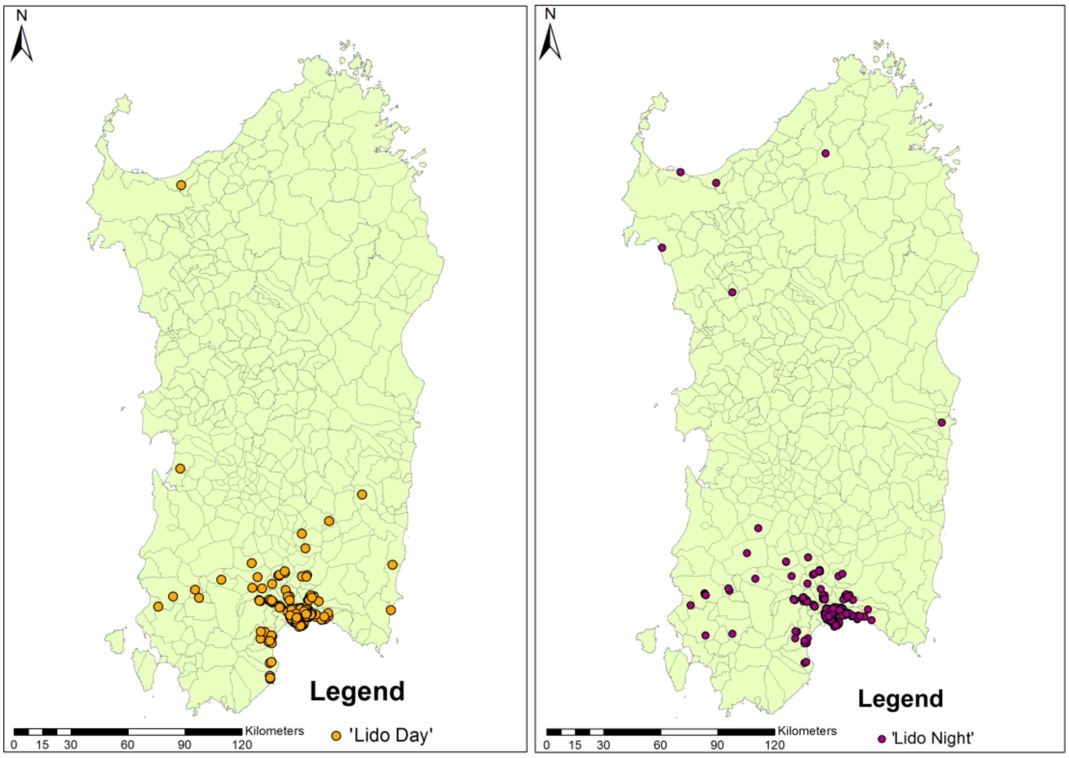


Figure 79. Provenance of the ‘Lido’ users during (A) ‘day’ period and (B) ‘night’ period.

In conclusion, this multi-dimensional analysis demonstrates that the ‘Lido’ area is popular among several and similar population groups independently from the temporal periods. This phenomenon may be explained thanks to the range of differentiated services offered by the place during the day and night, which are able to satisfy the requirements of customers. The most notable difference is the spatial clustering phenomenon of SMGI during the night, which particularly stresses the high attractiveness of the discotheque / bar services for customers during the night hours of the summer period.

7.8 Summary of results

The “Poetto” study offers an overview of the possible use of SMGI platforms to investigate what people observe, prefer and how they behave both in space and time in a specific public area. The main endeavor of the study was the operative application of the novel SMGI Analytics framework in order to test the assumptions about the analytical opportunities hailing from the approach for spatial planning. The results demonstrate that the SMGI Analytics framework may foster the elicitation of meaningful knowledge for urban and regional planning thanks to the novel tools and methods developed in this study, supporting the gain of both quantitative information and qualitative experiential knowledge concerning a specific geographic area.

The results obtained from ‘Poetto beach and Molentargius Park’ area investigation showed since the first methodological step, namely the data collection, the notable abundance of the Instagram data. This study takes advantage of this phenomenon in the research for investigating the users’ behaviors, preferences and opinions by means of analyses carried out on the spatial, temporal and user dimension of Instagram SMGI datasets.

The temporal analyses showed similar trends concerning the Instagram SMGI production and sharing during workdays and weekends, and identified the highest peaks of interest during the H: 2:00-3:00 PM period and the lowest ones during the period H: 5:00-6:00 AM for workdays and H: 07:00-08:00 AM for weekends. As claimed by Silva et al. (2013 A; 2013 B), the temporal patterns may be considered as a cultural footprint of the place, which may depict the main habits and behaviors of local communities.

Afterwards, the DB-SCAN clustering methodology, applied on the SMGI dataset, led toward the identification of the most popular areas for both the users sample and for each single individual. Different DB-SCAN parameters were used to identify the most suitable clustering solution, which was found to be ε = 5 meters (threshold distance) and *min_pts* = 10 (minimum number of features in a cluster). With the above parameters, the clustering analysis was able to detect over 200 clusters, which depict the higher attractiveness of the seaside area in respect to the inner area of this public space. Along the same vein, the modified clustering method, namely the FB-DBSCAN, was applied on the Instagram SMGI dataset to identify about 2.5K residential areas.

The geodemographic classification enabled the classification of the Sardinia territory at the census tracts level, relying upon official social-economic variables, and labels assignment to each identified group allowed the SMGI dataset' users profiling. The final census tracts classification into two hierarchical levels, demonstrated the capability to apply a method, commonly limited to commercial and private systems, in spatial planning analysis for gaining useful insights on local communities.

The user profiling results were used to develop a number of multi-dimensional analyses on the public space area in order to characterize the places and analyze the urban dynamics and preferences of specific users groups in space and time. First of all, the analyses demonstrated the capability to identify POIs in the area relying upon the detected clusters, the complementary SMGI extractions, conducted on Foursquare and Instagram Places, as well as, the integration of A-GI extracted from the Regional SDI. The results showed that the users' preference and movements are mainly attracted to the natural characteristic of the area, namely the Poetto beach, but also to several facilities related to leisure such as café, restaurant, bar, or to accommodation such as bed & breakfasts and hotels. Moreover, the identified POIs demonstrate the users movement in the area to attend the public hospital 'Ospedale Marino'.

Conversely, the results show a lack of attention by users towards the Molentargius Park area, which not presents any SMGI cluster. This phenomenon may be explained considering the structure of the park that offers opportunities to appreciate the natural attractions of the area by providing a number of services such as pedestrian and bicycle paths, as well as a number of sport activities and initiatives, which may be conducted during specific time periods of the year. On the one hand, these services are able to attract users, but probably due to the moving nature of the proposed activities, the users capability to take photo in the same or nearby geographic points is limited, dismissing the identification of clusters relying exclusively on points density. In this respect, the integration of SMGI from multiple social networks should help at improve the characterization of this area, identifying strengths and weaknesses according to users preferences and dynamics. Despite a number of potential limitations, this kind of analysis was able to explore in details the preferred users' places, and enabled to discover the most visited places and the social dynamics affecting the Poetto public space.

In a similar way, the second analysis investigated the preferences of a specific users group, namely the tourists, identifying which areas were the most appreciated by them, and, at the local scale, the most visited POIs, as elicited from the Foursquare SMGI. The identified POIs are all related to food and leisure, namely café, snack shops and ice cream shops, and the most popular areas are distributed on the first beach section, flanking the 'Marina Piccola' touristic harbor.

The third analysis dealt with the investigation of two nearby places within the Poetto beach. The analysis disclosed the dissimilarities between the two locations in terms of temporal patterns, users' profile typology and users' provenance. On the one hand, the results demonstrated that the 'Lido' area, providing several services to its customers related both to daily-life on the Poetto beach and to night-life, was able to strongly attract different user typologies during different temporal periods and also from distance places of Sardinia. On the other hand, the 'Free beach' area was visited by a low number of users and exclusively for limited time periods both during the year and also during the day. In addition, the users' provenance showed that the place was commonly visited exclusively by users residing in Cagliari or in the adjacent municipalities.

Finally, the fourth analysis concerned the investigation of a place, namely the 'Lido' area, for different temporal periods. On the one hand, the results demonstrated that the area is popular to the same user groups independently from the considered interval thanks to the differentiated set of offered services. On the other hand, the findings show how the 'Lido' is extensively used for its whole area during the day, probably denoting an use related to beach activities, meanwhile during the night period the users' contributions are clustered, revealing a high attractiveness level related to leisure services. This phenomenon demonstrates the particular nature of the place, which is able to change significantly its services nature in order to deal with the requirements of users that may differ from day to night.

7.9 Discussion

This chapter discusses the methodological approach adopted in the research towards a SMGI Analytics framework, through its application on a complex case study in order to gain useful insights about users dynamics and preferences for easing the governance and the decision-making processes. The different stages of the SMGI Analytics framework are analyzed in details and applied on an SMGI dataset extracted from the Instagram social networks and related to the public space of the Poetto beach and Molentargius Park in the municipality of Cagliari and Quartu Sant'Elena.

The proposed approach relies upon the different dimensions of the SMGI, which may disclose innovative analytical opportunities in spatial planning to investigate not only quantitative geographic facts, but also the preferences, the opinions, the dynamics and the behaviors of users concerning different places and localities. After the data collection, the methodology proposes the development of explorative spatial and temporal analyses aimed at identifying interesting patterns and dynamics within the study area. Afterwards, in order to take advantage of the user dimension, the chapter discusses the use of clustering techniques to detect the area of major interest in the area both for the whole users sample and then for each single user. Furthermore, the discussed SMGI Analytics methodology introduces the geodemographic classification of the Sardinia census tract, relying on the official census data made available by ISTAT. The geodemographic classification method builds on the use of 43 socio-economic variables in order to segment the regional territory in different groups at two different hierarchical level. The first level consists of 6 groups, while the second hierarchical level is composed by 18 sub-groups, which enable to distinguish the principal characteristics of resident people.

The results of the geodemographic classification are then processed and labeled according to the variables measurements for developing the profiling of Instagram SMGI dataset. This methodological step is fundamental for gaining useful insights about the behaviors of users, as well as, to investigate in details the different patterns and dynamics affecting the study area. From an operational perspective, the discussed approach aims at characterizing the Poetto Beach and Molentargius Park areas according to the spatial distribution of SMGI contributions, stressing the major differences that may emerge also between bordering places in the same study area according to the preferences of several users groups. It is possible to assume that a different SMGI density may be considered as an indicator of the area attractiveness and should be used for developing further focused analyses. As shown in the Iglesias case study, a high concentration of SMGI in certain areas was representative of the users major interest towards specific locations, that were the main public spaces of the municipality. In the Poetto case study, in spite of an analysis entirely conducted within a public area, the high density of points may be used to both identify the specific locations or POIs attracting the users and investigate the nature of their services offers. In this way, the understanding of the reasons behind the users dynamics may be used as advantage for developing policies oriented to the real users requirements, fostering the success of such public initiatives.

In literature, several novel approaches for the use of VGI and SMGI are proposed, but no one to our knowledge, at the current time, proposes the formalization of a methodology such as the SMGI Analytics framework, to investigate the social media data from different perspectives and at different geographic scales. Often, the proposed analyses are limited to the investigation of a specific dimension or exclusively for a specific purpose, dismissing the great knowledge potential enclosed in this information. Moreover, the user dimension is commonly not considered, limiting de facto the opportunities to explore the behaviors and the dynamics of different users groups insisting on a geographic area.

In this respect, the SMGI Analytics exploits the capabilities and the potentialities of the geodemographic classification approaches, usually developed for commercial purposes, in order to consider and analyze the SMGI user dimension too. The results provide insights on users preferences and dynamics, which would not be available through other tradition data sources used in urban and regional planning.

CHAPTER 8

CONCLUSIONS

8.1 Introduction

This chapter summarizes the main key concepts and findings of the study, providing several considerations on the base of the achieved results. The chapter is organized as follows. The next section concerns the implications of the research questions, presented in the Chapter 1, discussing the findings and the main results. The last section concludes the manuscript and suggests potential future research agenda.

8.2 Main research results and findings

Since the last decade, continuous advances in the ICTs, the Internet, and more recently, the Web 2.0 technologies are increasingly channeling digital GI into the users' daily life, causing a paradigmatic shift in GI production, dissemination and consumption. In addition to the A-GI, namely the official information produced by public authorities, the VGI and SMGI, as voluntary contributed and shared GI, are fostering the development of bottom-up initiatives by the users, as well as, disclosing innovative analytical scenarios for urban and regional planning practices (see Chapter 1).

After an analysis of the spatial planning evolution, including a focus on the role played by participation, inclusivity and experiential knowledge into planning processes (see Chapter 2), the discussion debates the main opportunities and issues arising from the different nature of the digital GI sources (see Chapter 3).

Beginning from the literature review, the thesis addresses the different research questions:

1. What is the nature of SMGI?
2. What instruments are required to exploit SMGI in practices?
3. What analyses are required to take advantage of A-GI and SMGI?
4. What geographic scales are the most appropriate for using SMGI in practices?
5. How may the SMGI generated knowledge be used for supporting spatial planning?

8.2.1 The SMGI nature and the instruments for exploiting this information in practice

The nature of SMGI and the instruments that may be required to successful use this kind of information for supporting design, analysis and decision-making in spatial planning practice, are investigated by assessing several theoretical and methodological approaches currently used both in research and practices. Indeed, a review of existent studies demonstrates that different approaches and methodologies are proposed in order to take advantage of VGI and SMGI in urban and regional planning. Currently, despite this information suffers from heterogeneity and from incomplete coverage, it plays a major role in exploratory analysis and in integrating official datasets (see Chapter 4). However, from an analytical perspective, many approaches are based exclusively on the analysis of one or two SMGI dimensions, commonly extracted from only a social media platform, without evaluating the integration of this information with other sources (Purves and Derungs, 2015; Huang et al. 2013).

Moreover, SMGI offers a particular data structure that requires the use of suitable tools and analytical methods to properly manage and elicit useful knowledge for practices. Several authors suggest the development of integrated analysis on multiple dimensions (Campagna et al. 2013), or coupling multiple SMGI sources (Ostermann et al., 2015; Campagna et al. 2015), in order to evaluate and investigate more effectively the nature of this geographic information. In the light of these considerations, the thesis proposes the development of user-friendly tools in order to ease the collection, the management and the analysis of SMGI (see Section 5.3).

The thesis introduces the SPATEXT suite, which is a collection of tools developed to ease the extraction and the management of SMGI retrieved from multiple social networks and the contextual integration of this information in GIS environment for analysis. The SPATEXT suite tools are developed with the aim of facilitating the access and the use of social media data to planners offering ad-hoc designed functionalities. The SPATEXT suite functionalities are evaluated in practice through several case studies conducted at different geographic scales using SMGI extracted from different sources, as described in Chapter 6 and Chapter 7.

The SPATEXT suite evaluation demonstrates that the ad-hoc designed and developed user-friendly tools are able to deal with the hurdles regarding the access, the management and the analysis of SMGI, directly in GIS environment, fostering the development and the application of the SMGI Analytics framework. The results obtained from the different case studies demonstrate the fit-to-purpose of the several developed instruments and their capability to ease the collection, the management and analysis of SMGI. These tools may complement the traditional tools used by planners, offering opportunities for developing and applying innovative analytics framework, as well as for integrating the availability of GI with the novel digital sources, avoiding programming efforts to practitioners.

8.2.2 The analyses for using SMGI and A-GI at different geographic scales

After the introduction of the SPATEXT suite, the thesis discusses the development of the SMGI Analytics. The SMGI Analytics is a novel analytical framework that fosters the integration of A-GI and SMGI in a GIS environment, including several explorative, qualitative and quantitative analyses in order to proficiently elicit knowledge from SMGI for spatial planning analysis and decision-making (see Chapter 5). The framework consists of several stages, which shape the operational workflow to conduct multi-dimensional analyses on different topics and at different geographic scales. The findings provide insights concerning users' opinions, requirements, preferences, behaviors and urban dynamics, enabling the investigation and characterization of places by means of information which is not available through traditional data sources, commonly used in planning practices.

The proposed methodology is used to provide information about spatial and temporal patterns, popular POIs, users preferences and opinions, land uses and urban dynamics, as well as, to gain information about user groups interests and concerns through the development of an user profiling approach. Several case studies, conducted at different geographic scales using SMGI collected from different social networks, are developed in order to investigate local communities' perceptions on relevant topics for spatial planning and the geography of places (see Chapter 6).

The early experience was conducted on SMGI extracted from Twitter in order to analyze, at the global scale, the users opinions about the cyclone Cleopatra occurred in Sardinia (Italy) in November 2013. The results are affected by the Twitter Search APIs limitations for data collection, concerning both the reduced availability of data (1% of the total volume) and the capacity to access SMGI posted at most one week before the extraction date, which provoke several hurdles for gaining insights from SMGI analyses. Despite the identified limitations, the study was able to discover the growing trend of the topic among the social network users worldwide and particularly the major role played by breaking news Twitter accounts for disseminating information concerning the cyclone. The high retraced re-tweet phenomenon among users demonstrates the powerful capacity of certain Twitter accounts to lead the discussion on the social network also in the case of a topic related to a specific and no popular area as Sardinia. The spatial patterns of contributions expose the major interest of Italian users, in respect to other ones, to the topic, raising interesting questions about the direct relationships between interest and spatial proximity. Finally, the textual analyses results demonstrate the characterization of the discussion according to the geographic area. Textual tags identified in Italy, in spite of several similarities with the worldwide tags, present a number of differences, which are provoked by the geographic context of the social network users. This finding confirms the notable opportunities that textual analyses may disclose to elicit useful information from SMGI sources.

Secondly, the investigation of landscape perception by users through the analysis of YouTube SMGI exposes an almost homogeneous distribution of videos across the Sardinia Region, but at the Provincial level, the results display a number of differences. Indeed, the findings show a different perception of landscape by users according to different geographic zone. The SMGI related to the Province of Nuoro shows a high vocation toward the inner areas landscape, meanwhile the Province of Cagliari and Olbia-Tempio videos perceive the landscape as strictly related to the coastal areas. This result may be due to the specific characteristics of the geographic areas, which present different natural and cultural features, and may offer different services and attractions to users and visitors, leading toward the recognition of the landscape in different ways strongly affected by territorial socio-cultural biases.

Similarly, the investigation of neighborhoods perception at the local scale in the municipality of Cagliari demonstrates the SMGI capability to highlight in a bunch of words the main characteristics, and the different perception of each examined area. On the one hand, the ‘Castello’ and the ‘Marina’ neighborhoods are depicted by means of characteristics concerning the architecture, the events and the inherent area functionalities. In this case, the historical value of the neighborhoods is strongly represented by the video titles and captions. On the other hand, the ‘Is Mirrionis’ neighborhood and the ‘Molentargius Park’ are depicted exclusively by means of toponyms and activities. As a matter of fact, the ‘Is Mirrionis’ is a poor dwelling area built during the 1960s city sprawl, but this characterization is not provided in texts, while major emphasis is given to the sport facilities provided by the place. The ‘Molentargius Park’, probably due to its specific function in the urban environment, is described through its flora and fauna principal features, allowing the detection of recreational and park functionalities, immediately. Furthermore, the results concerning the overall opinion of the neighborhoods are particularly interesting because of the similarity with the findings of a study conducted through the platform Place, I Care! (PIC) and Cagliari, I Care! (CIC) (Campagna et al., 2013). This phenomenon raises interesting questions about the invariance and reliability of SMGI for spatial analysis, inasmuch the studies deal with different users and time periods, and should be further investigated in future studies.

The geography of the places’ investigation, carried out through Instagram SMGI related to the Iglesias municipality in Sardinia (Italy), reveals the SMGI capability to analyze the users’ preferences and behaviors in space and time. In addition, the findings provide interesting insights about the use of Instagram by the involved users. The social platform is mainly used in the built environment, with approximately the 89% of the contents taken in residential or commercial and service areas. Moreover, sheer volume of data are produced during spring and summer and in particular during two peaks of interest detected in the daily trend, that are 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. These findings suggest an increased use of the social network during the meals or breaks times. Indeed, Instagram is a mobile photo and video sharing social network which enables users to upload photographs and short videos, explore other users' feeds and use the geotag functionalities to georeference images. The service is gaining increasing popularity and, since 2012, it offers users a web profile, showing a selection of personal recently shared photographs, personal information, and other details. In addition, the explore functionality and the filters option for modifying photos extend the range of the social network capabilities. On the one hand, the Instagram functionalities are able to involve users in selecting and sharing personalized photos, but on the other hand, the use of the social network is limited to certain time periods, when users have free time or leisure in order to explore other friends photos or to update their own profiles. Therefore, the temporal preferences of users, who contribute to the social network, may differ accordingly to different geographic areas and this phenomenon should be considered during investigation, especially during the identification of temporal patterns and peaks of interest. As a matter of fact, the identification of a peak of interest might be the result of an increased social activity by users due to specific temporal periods and not the increased interest toward a specific topic or geographic area. Nevertheless, the case study findings expose a predictable pattern for the users' behavior in contributing to Instagram, easing the identification of real spatial-temporal phenomenon.

In conclusion, all the methodological approaches applied in the discussed case studies are based upon the choice of an appropriate geographic scale, in order to obtain useful knowledge from SMGI for spatial planning. However, the geographic scale required for analyses is strongly affected by the kind of pursued investigation, as well as, by the specific social network used for the data collection. Particularly, the social networks may differ in terms of diffusion among users worldwide, SMGI data model and debated topics in shared information, resulting in the necessity to vary the geographic scale, accordingly. As a matter of fact, the exemplificative case studies elicit knowledge from SMGI by relying upon the global, regional or local scale, due to the different social networks used for data collection.

8.2.3 The SMGI generated knowledge for spatial planning

Shifting the discussion from the patterns of social network utilization towards the opportunities for spatial planning, the findings demonstrate that SMGI may be used to identify both the most appreciated and popular places in the municipality, as well as, to detect different land uses by temporal and spatial patterns thanks to the development of several clustering analyses.

The Iglesias case study suggests that the locations, attracting the major users' interest within the Iglesias municipality, are placed in the public spaces of the city center and in the historical areas. The most appreciated places are the historic Cathedral of 'Santa Chiara', 'Via Matteotti', namely the main avenue for leisure and night life of the municipality, the squares 'Piazza La Marmora' and 'Piazza Sella', and the urban public garden. These results may be considered an important descriptor of the city physical structure, as well as, a valuable indicator of users' preferences and dynamics. The presence of historic area near the city center, coupled with a number of public spaces and services might stoke the users to frequent these areas. On the other hand, the results of land use investigation, carried out by clustering analysis, allowed the identification of users residential areas and the identification of several buildings un-mapped in official information. These findings demonstrate the profitable use of SMGI in respect of the questionnaires, which should be conducted on a local community's sample for obtaining similar results.

The early experiences' results informed the development of a more complex case study on the public space area of 'Poetto beach and Molentargius Park' in the municipality of Cagliari in Sardinia. The multi-dimensional analyses conducted on a one year Instagram SMGI dataset reveal the high attractiveness of the seaside area along the entire 'Poetto' beach, in respect to the 'Molentargius Park' area, which in turn shows a notably lower volume of data (See Chapter 7).

The results of temporal analyses reveal not surprisingly that the highest volume of contributions occurs during the summer, denoting the particular nature and the major use of this area for activities mainly related to seaside activities. At the same time, the users appear to prefer the weekends to visit the place, probably due to the increased availability of free time and the opportunities to attend leisure activities. In details, the kernel density results depict as the most popular areas the locations situated near the touristic harbor 'Marina Piccola' and in the initial part of the 'Poetto Beach'.

Beyond the development of explorative spatial and temporal analyses, the research developed a set of multi-dimensional analyses for investigating:

- the different attractiveness of locations in the area through clusters;
- the POIs and the services contained in the clusters that attract users;
- the different user groups emerging from the geodemographic classification and their behaviors;
- the different utilization of nearby areas presenting strong different natures;
- the different uses of the same place during diverse time periods.

First of all, the results show that the public space area is mainly visited for the presence of both the beach and a number of leisure amenities, namely café, restaurant and bar, thanks to the integration of multiple SMGI sources, which enable to identify the most visited POIs in the area.

Secondly, the geodemographic classification exposes the potentiality to partition the Sardinia territory in different population groups according to socio-economic variables, able to depict the main population characteristics. The study proposes a two-tier geodemographic classification, consisting of 6 groups for the first hierarchical level and 18 sub-groups for the second one. The second hierarchical level is used for developing the final user profiling, enabling a set of focused analyses on the preferences of specific user groups in the study area. Hence, the preferences expressed by the tourists group lead toward the identification of the most frequented places by this category along the 'Poetto Beach', such as café, snack shop and ice cream shop. The concentrated tourists' SMGI distribution in the beginning of the beach may be explained considering the public transport routes. Indeed, the public transport service, originating from the Cagliari inner metropolitan area, proceeds along the 'Poetto beach' starting from the identified area. Furthermore, this area offers a number of services which appear to satisfy the tourists' requirements both in terms of seaside activities, leisure and relax.

More interestingly, the results obtained by comparing the 'Lido' area with the 'Free beach' one reveal the different degree of attractiveness offered by the two locations to the different user groups. The 'Lido' area, due to its multi-purpose nature, offers several service typologies able to meet the customers' expectation all year long. On the other hand, the 'Free beach' area is lived exclusively during the summer, denoting the lack of services able to attract the users' interest beyond the seaside offer. This lack of services is also evident when evaluating the users' provenance, the daily distribution of SMGI and the number and typology of users frequenting the place. The obtained results reveal that the services supply represents an important factor to increase the attractiveness of a location to users. This is mainly evident by comparing the two areas. While on the hand, the 'Lido' and 'Free beach' areas share the same natural tangible resources (i.e. Poetto beach, natural attraction, cultural attraction, landscape) because of their spatial proximity, on the other hand, they expose notable differences in users participation due to the completely differences supply of services. In terms of policies, these results may suggest the development of strategies oriented toward the services supply, in order to satisfy the requirements of users who currently or in future may be interested to visit the area.

The analysis of the 'Lido' area, conducted for both the 'day' and the 'night' periods, demonstrates that the area is popular to the same user groups, independently from the temporal periods, probably thanks to the wide range of offered services. Nevertheless, the findings show a SMGI spatial clustering phenomenon during the 'night' period, which may be directly connected to the presence of specific leisure services, active during that time band. A direct investigation shows that the SMGI clustering is mainly related to the 'Lido' discotheque area and its entrance. This kind of result, obtained coupling the spatial, temporal and user dimensions, may be used to investigate not only the most attractive places of an area, but also the reasons behind the attractiveness and the different typologies of user who attend these places according to diverse time bands.

From an analytical perspective, the methodological approach proposed in the thesis might be proficiently used in planning practices, easing to take into account the real users' requirements and preferences, as depicted by their behaviors and movements freely provided through the social networks. The lesson learnt from this kind of analyses may be used in spatial governance at the regional and at the local level in order to identify the factors which may help in developing sustainable and successful policies to act in less developed areas with focused interventions. Indeed, the adoption of an integrated framework, exploiting both technical and experiential knowledge supplied by A-GI and SMGI, respectively, may help in identifying the reasons behind users' preferences and urban dynamics. In this respect, the findings provide an overview of potential SMGI use 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.

The findings might be used in several models of the Steinitz Geodesign framework. The results of explorative spatial distribution, as shown in the Iglesias and Poetto case studies, may be used as input datasets in the representation models, helping to depict the current utilization rates of certain areas and allowing the identification of the most used and less used locations within the geographic context. At the same time, the cluster analyses results may be used to identify the most appreciated POIs and the nature of the supplied services, enriching the knowledge base for further analysis. Similarly, the temporal analysis findings, eliciting the users dynamics and movements, may be used to calibrate specific models that simulate the social processes affecting the area. This way, official information and SMGI may help to investigate the evolution of phenomena in space and time, feeding predictive models, which may help the involved actors in evaluating the current state of the geographic context and potentially the future trends.

Shifting the discussion to the evaluation stage, the findings may be used for gaining insights on the one hand about the positive factors of the study area, as well as, on the other hand, about the weaknesses, which affect the less visited area. This opportunity arises for example from the investigation of supplied amenities in different locations within the study area, as demonstrated by the Poetto and Molentargius public space analysis. In addition, the identification of the potential causes provoking a lack of utilization, might be used to develop public policies and strategies oriented at solve these issues, such as the shortage of specific services concerning sport, leisure, and cultural activities, the lack of transport routes during certain temporal periods, or the presence of environmental and public safety concerns affecting specific locations.

In turn, the identification of success factors, as depicted by the clusters of interest, may be used in the impact model stage by suggesting which components should be also evaluated in order to assess the alternative solutions proposed during the geodesign study. SMGI may help for taking into account the user requirements and for assessing the proposed scenarios with real-time information and by means of direct observations of the social dynamics. Finally, the findings may be used by decision-makers in the decision stage in order to discriminate among the different alternatives, paying specific attention to the concerns of users and selecting among the solutions the one which may implement strategies able to satisfy the requirements of the involved stakeholders.

The municipal authorities may rely upon this kind of analysis for developing a set of public policies and strategies. The findings obtained from the different components of SMGI, namely spatial, temporal and user, may be used in order to value certain municipal areas assessing the supply of social services and amenities able to meet the users requirements. Indeed, the obtained results are directly connected with the intertwined systems that define a ‘smart city’, namely the infrastructural, the managerial and the social (Nam et Pardo, 2011). These systems may be further partitioned into a set of more specific factors (Chourabi et al., 2012), which should dynamically collaborate and support each other to guarantee the ‘smart’ strategies success.

Therefore, information regarding the SMGI spatial distribution may be used to implement strategies concerning the development of urban services such as pedestrian and bicycle paths, which may increase the utilization and appreciation of less used locations. For example, the final section of the Poetto beach shows a lower utilization rate than the initial segment, and this issue might be addressed developing this kind of urban services in order to attract more people to visit the areas.

The increased appeal of certain services that supply sport, cultural or leisure amenities, as identified by a number of identified clusters in the Poetto beach, may be used as a guide for driving the development of similar or complementary services in other public space locations.

In this respect, the identification of common movement dynamics among users may be used for the design of pedestrian routes able to connect different public spaces. This kind of route is already existent, connecting for example the Molentargius Park with the Poetto beach, but the findings show a shortage in the use of this path. Hence, the lack of clusters in the park may suggest the development of areas equipped for sports and leisure, which should foster an increased participation of users if coupled with the supply of services and amenities in compliance with the environmental requirements of the area (i.e. bird watching, sports competition, training activities, drink & food).

The different density of SMGI contributions, if evaluated in real-time, may allow the investigation of users presence and these data may be used as human sensors to take (near) real-time decisions. For example, the increased presence of visitors due to an event may be elicited from a real-time SMGI analysis and information about potential peaks of interest may be used for directing control activities ensuring the public safety.

The major attractiveness of certain locations, due to the specific natural characteristics, may be inferred from the increased density of SMGI contributions, as well as, from the textual analysis of contents. This kind of analysis may help at identifying significant placemarks related to natural, landscape or cultural characteristics of the area, which may be used for leading the development of strategies oriented at valuing the area for touristic or cultural purposes, as well as, for identifying areas suitable to be pedestrianized. This is the case of the neighborhoods investigation that stressed the principal characteristics and main placemarks attracting the users attention, fostering the identification of the characteristics most suitable to be considered in a walkability analysis.

Similarly, the temporal distribution may be used for informing public policies concerning the timetable of local communities. This information may represent a valuable data for the reinforcement of urban transportation routes. The findings show the increased use of the Poetto beach and Molentargius area during weekends and summer. Hence, a number of policies should be developed in order to ensure the increased presence of public transports during these periods in order to limit the use of private vehicles.
At the time being, the municipal authority is highlighting the need to limit the use of private transports in the area and is fostering the development, as well as the use of pedestrian routes and bicycle lanes. In addition, the public transportation offer is strengthen during the daily summer period in order to guarantee a direct connection from several city neighborhoods to the Poetto beach and vice versa. Nevertheless, the findings demonstrate that the area is also used during night for several leisure amenities supplied by a number of private locals. These locals are able to attract a notable volume of users, but the lack of night public transportation routes compels the use of private vehicles, which may provoke negative effects both on the novel predicted pedestrian nature of the beach and on the environmental system. Therefore, the local authority should consider the obtained results for developing a number of night connections between the city and the Poetto area in order to satisfy the users requirements. The findings of temporal trends might be used to establish proper timetables able to address the locals requirements and to avoid excessive costs for the municipality.
The integration of information concerning spatial and temporal patterns might be used by the local authority to identify the areas less visited by the users and to investigate the underlying causes. A textual analysis might for example depict a less level of safety for certain areas due to reasons concerning crimes and illegal activities. These results should be used to address the depicted issues through increased controls in order to ensure the public safety.
Finally, the temporal patterns may be used by the local authority to implement urban time policies, namely policies that intervene in the time schedules and time organization in order to regulate the human relationships within the city (Mareggi, 2002). A direct action on the services timetables may extend the availability of public open spaces and enable the design or the rethink of strategies concerning the mobility, leading a number of actions which may deal with the demand of transportation exclusively in certain time periods (Mareggi, ibidem). The temporal analysis findings enable to introduce the time dimension in the decision-making process, extending the opportunities for local authority to address the different requirements of stakeholders and local communities with reliable solutions (Dente, 1997) built upon the novel SMGI knowledge that traditional territorial government usually is not able to take as an advantage.
The temporal trends show an increased used of the public spaces during specific times; therefore it is possible to argue that a set of policies oriented toward both the increased mobility and the development of public initiatives in these time bands should be a suitable solution to meet the users requirements.

Moreover, the increased attention to time may help in depicting the ways of using places by residents and temporary inhabitants, as well as, the rhythms of individuals' life in the city (Bonfiglioli, 1997).

The user dimension, as depicted by the geodemographic results may be used for developing strategies oriented at guaranteeing the satisfaction of people on the base of groups preferences, as well as, for identifying the causes behind the not utilization of a certain areas.

The results show that the tourists prefer to visit in particular the first segment of the Poetto beach, while their presence is strongly limited both in the other beach parts and in the Molentargius Park. This phenomenon might be addressed by the local authority strengthening the promotion of the whole area and by developing a set of social initiatives and mobility interventions, which may increase the appeal of the less visited area. This issue should be addressed for example proposing a number of cultural routes or a number of daily and night events which value the whole beach area.

In addition, the insights about the dynamics and places preferences expressed by the users and elicited by the geodemographics groups may be used for identifying the common behavioral patterns and the most distinctive characteristic of each group. In the light of the obtained results, it is possible to calibrate a number of policies, which take into account the satisfaction of as many groups as possible in order to address their requirements.

Moreover, the geodemographic results might be used by the local authority for developing a number of analyses at the urban level investigating the residential patterns of groups and potentially developing a set of urban time policies informed by the different groups' timetables, dynamics and urban structure. The results may be integrated with other official information for developing a set of differentiated services according to the patterns and preferences of identified groups. For example, the SMGI insights might be used for distinguishing between tourists concerned on the natural and cultural attractions and the tourists concerned to the beach and the leisure activities, supplying a set of differentiated suggested routes and paths across the city and the public spaces.

At the same time, this kind of information related to residential users groups and their preferred timetables may be used to calibrate the open and close times of certain public services, as well as, to avoid the traffic congestion in the municipality. The results of temporal peaks and spatial distribution may be coupled in order to identify the places more affected by congestion issues, and several policies may be implemented to avoid this kind of problem.

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 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 their 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.

8.3 Conclusions

The findings obtained through the SMGI Analytics framework contribute to demonstrate the possible use of SMGI platforms to investigate what people observe, evaluate, and how they behave both in space and time. The primary goals of the study, namely the formalization of a novel SMGI Analytics methods and the development of tools, may help to access different social network data sources and extract meaningful knowledge for spatial planning.

The study proposes an approach that may give several empirical contributions to the evaluation of social media data as sources for analysis in the spatial planning domain. First of all, the formalization of the SMGI Analytics and its application propose a replicable methodology for taking advantage of experiential and voluntary geographic information, coupled with official information, in order to investigate different geographic areas in space and time. Secondly, the methodological approach demonstrates how the integration of A-GI with SMGI may pave the way for the development of multi-dimensional analyses able to take into account different dimensions usually neglected in traditional spatial planning practices. Several spatial analyses and techniques are supplied in order to demonstrate how SMGI may be directly used and integrated with traditional authoritative spatial data layers in GIS environment.

Furthermore, the obtained results show the opportunities of this novel method to collect updated information about places, dynamics and user behaviors, in order to deal with the hurdles of traditional methods, such as on the field surveys and direct observations, usually limited by time and cost constraints.

The experiential knowledge, which may be elicited from the SMGI, could be used in the novel theoretical approach of the Geodesign and in the SEA methodological approach in order to supply a deep knowledge of the geographic context and the cultural and social dynamics for enriching the development of sustainable processes. Similarly, the ‘smart city’ initiatives may gain value from the broader and pluralist knowledge of the places enclosed in SMGI in order to develop advanced technological solutions, which integrate official and experiential information with sensor data infrastructure, fostering the implementation of strategies informed by local communities in a bottom-up approach. Indeed, the SMGI opportunities might foster such scenarios where a city planner is able to include the local community’s concerns and interact with them to design alternative projects and select future development options by a constructive and participative dialogue about places, eventually.

Nonetheless, it is important to be aware that the SMGI datasets should be not considered representative of the whole population. In fact, 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 active groups highly affect the phenomena under observation in SMGI. In the future, a wider diffusion may occur to this respect as suggested by the current social network growth trends, but since the time being, more research is needed to assess the full potential of SMGI and several issues should be addressed regarding data quality and representativeness. However, meanwhile these issues should be further investigated, the potential of these new data sources already far exceeds those of more traditional inquiring methods and tools used in urban and regional planning.

In conclusion, the methodological approach proposed in this thesis, may open new analytical scenarios for planners, as well as new research challenges, which aim to use A-GI and SMGI for gaining pluralist knowledge and developing users-oriented policy making in spatial planning and governance. The SMGI Analytics application moves a step forward in this direction, proposing a number of methods and analyses for easing the design or selection of plan options and fostering the direct observation and identification of strengths and weaknesses in an area by means of updated, accessible and potentially unbounded sources of information such as the social networks. To what extent SMGI may contribute to enhance the quality of knowledge and eventually bring innovation to spatial planning represents a very challenging and stimulating opportunity for future research.

REFERENCES

100 Social Networking Statistics & Facts (2012). *100 Social Networking Statistics & Facts For 2012.* Retrieved from: http://visual.ly/100-social-networking-statistics-facts-2012 [Accessed 2013 Nov 13]

Adam, L. (1996). Electronic communications technology and development of Internet in Africa. *Information Technology for Development*, 7(3), 133-144.

Adnan, M., Longley, P. A., Singleton, A. D., & Brunsdon, C. (2010). Towards Real-Time Geodemographics: Clustering Algorithm Performance for Large Multidimensional Spatial Databases. *Transactions in GIS*, 14(3), 283-297.

Al Zamal, F., Liu, W., & Ruths, D. (2012). Homophily and Latent Attribute Inference: Inferring Latent Attributes of Twitter Users from Neighbors. *In Sixth International AAAI Conference on Weblogs and Social Media ICWSM*.

Al-Bakri, M., Fairbairn, D., 2012. Assessing similarity matching for possible integration of feature classifications of geospatial data from official and informal sources. *International Journal of Geographical Information Science*, 26(8), pp. 1437-1456.

Andrienko, G. L., & Andrienko, N. V. (1999). Interactive maps for visual data exploration. *International Journal of Geographical Information Science*, 13(4), 355-374.

Annoni, A., and M. Craglia. 2005. Towards a Directive establishing an infrastructure for spatial information in Europe (INSPIRE). *In Proceedings of GSDI-8 from Pharaohs to Geoinformatics: The role of SDIs in an information society*, Cairo, April 16–25. Retrieved from: http://gsdidocs.org/gsdiconf/GSDI-8/papers/ts_47/ts47_01_annoni_graglia.pdf [Accessed 2014 March 10]

Antoniou, V. (2011). *User generated spatial content: an analysis of the phenomenon and its challenges for mapping agencies* (Doctoral dissertation, UCL (University College London)). Retrieved from: http://discovery.ucl.ac.uk/1318053/ [Accessed 2014 March 18]

Antony, S. (2010). *Computational social science* (Doctoral dissertation, Cochin University Of Science And Technology). Retrieved from: http://dspace.cusat.ac.in/jspui/bitstream/123456789/260/1/Siji_CompSocSci.pdf [Accessed 2015 October 11]

Arnstein, S. R. (1969). A ladder of citizen participation. *Journal of the American Institute of planners*, 35(4), 216-224.

Arthur, D., & Vassilvitskii, S. (2007). k-means++: The advantages of careful seeding. *In Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms* (pp. 1027-1035). Society for Industrial and Applied Mathematics.

Artz, M. , Dangermond, J. , Ball, M. , Abukhater, A. (2010). *Changing Geography by Design*. Selected readings in GeoDesign, Esri Press, Redlands.

Assunção, R. M., Neves, M. C., Câmara, G., & da Costa Freitas, C. (2006). Efficient regionalization techniques for socio-economic geographical units using minimum spanning trees. *International Journal of Geographical Information Science*, 20(7), 797-811.

Ball, M. (2010). *What's the distinction between crowdsourcing, volunteered geographic information and authoritative data. Spatial Sustain*. Retrieved from: http://www.sensysmag.com/spatialsustain/whats-the-distinction-between-crowdsourcing-volunteered-geographic-information-and-authoritative-data.html [Accessed 2015 Jan 15].

Batty, M., Axhausen, K. W., Giannotti, F., Pozdnoukhov, A., Bazzani, A., Wachowicz, M., Ouzounis, G. & Portugali, Y. (2012). Smart cities of the future. *The European Physical Journal Special Topics*, 214(1), 481-518.

Bégin, D., Devillers, R., & Roche, S. (2013, May). Assessing volunteered geographic information (VGI) quality based on contributors' mapping behaviours. *In Proceedings of the 8th international symposium on spatial data quality ISSDQ* (pp. 149-154).

Berry, M. W., & Kogan, J. (Eds.). (2010). *Text mining: applications and theory*. John Wiley & Sons.

Bertrand, F. (2012). *Visualizing the Twitter social graph, Part 1: Getting the data. In: Technical write-up on Recollect Engineering*. Retrieved from: http://code.recollect.com/post/20476037331/visualizing-twitter-social-graph-pt1 [Accessed 2013 Nov 13]

Beyer, K., Goldstein, J., Ramakrishnan, R., & Shaft, U. (1999). When is "nearest neighbor" meaningful?. In Database Theory—ICDT'99 (pp. 217-235). Springer Berlin Heidelberg. LNCS 1540: 217–235. doi:10.1007/3-540-49257-7_15. ISBN 978-3-540-65452-0.

Birkin, M., Clarke, G., & Clarke, M. P. (2002). *Retail geography and intelligent network planning*. John Wiley & Sons.

Bishr, M., & Mantelas, L. (2008). A trust and reputation model for filtering and classifying knowledge about urban growth. *GeoJournal*, 72(3-4), 229-237.

Bonfiglioli, S. (1997). Che cos'è un cronotopo. *In Bonfiglioli, S., & Mareggi, M. (Eds.) Il tempo e la città fra natura e storia: atlante di progetti sui tempi della città*. Istituto Nazionale di Urbanistica. Urbanistica quaderni, n. 12, pp.90-92 (Roma, Inu Edizioni).

Bruns, A. (2006). Towards produsage: Futures for user-led content production. 275-284. *In Sudweeks*, *Fay, Hrachovec, Herbert, & Ess, Charles (Eds.)* Cultural Attitudes towards Communication and Technology 2006, 28 June - 1 July, Tartu, Estonia

Budhathoki, N. R., & Nedovic-Budic, Z. (2008). Reconceptualizing the role of the user of spatial data infrastructure. *GeoJournal*, 72(3-4), 149-160.

CACI. (2014). *The Acorn user guide. The consumer classification*. Retrieved from: http://acorn.caci.co.uk/downloads/Acorn-User-guide.pdf [Accessed 2015 September 12]

Caliński, T., & Harabasz, J. (1974). A dendrite method for cluster analysis. *Communications in Statistics-theory and Methods*, 3(1), 1-27.

Callcredit Information Group Ltd (2015). *The next generation of segmentation. Cameo UK*. Retrieved from: https://www.cameodynamic.com/media/8654/cameo-uk.pdf [Accessed 2015 September 13]

Callingham, M. (2003). Current commercial sector use of geodemographics and the implications for the ONS area classification systems. *Personal Communication*. In Vickers, D., & Rees, P. (2007). Creating the UK National Statistics 2001 output area classification. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 170(2), 379-403.

Campagna, M. (2005). *GIS for Sustainable Development*. In: Campagna, M. Eds. GIS for Sustainable Development, pp. 3–20. CRC Press, Taylor & Francis Group, Boca Raton, USA (2005)

Campagna, M. (2014 A). The geographic turn in Social Media: opportunities for spatial planning and Geodesign. *In Computational Science and Its Applications–ICCSA 2014* (pp. 598-610). Springer International Publishing. DOI: http://dx.doi.org/10.1007/978-3-319-09129-7_43

Campagna, M. (2014 B). Geodesign from Theory to Practice: From Metaplanning to 2nd Generation of Planning Support Systems. *Tema. Journal of Land Use, Mobility and Environment.*

Campagna, M (forthcoming D). Why social is special when it goes spatial?. *In ENERGIC book (forthcoming)*.

Campagna, M., & Craglia, M. (2012). The socioeconomic impact of the spatial data infrastructure of Lombardy. *Environment and Planning-Part B*, 39(6), 1069.

Campagna, M., Kudinov, A., Ivanov, K., Falqui, R., & Anastacia, G. (2013). Place I care! Crowdsourcing planning information. *In AESOP-ACSP Joint Congress*. Dublin.

Campagna, M., Floris, R., Massa, P., Girsheva, A., & Ivanov, K. (2015). The Role of Social Media Geographic Information (SMGI) in Spatial Planning. *In Planning Support Systems and Smart Cities* (pp. 41-60). Springer International Publishing.

Caragliu, A., Del Bo, C., & Nijkamp, P. (2011). Smart cities in Europe. *Journal of urban technology*, 18(2), 65-82.

Castelein, W., Grus, Ł., Crompvoets, J., Bregt, A. (2010). A characterization of volunteered geographic information. *In Proceedings of the 13th AGILE International Conference on Geographic Information Science*, Guimarães, Portugal, pp. 32–39.

Caverlee, J. (2010). A few thoughts on the computational perspective. *Presented during the specialist meeting on spatio-temporal constraints on social networks.* 13–14 December 2010, Santa Barbara, CA.

Chourabi, H., Nam, T., Walker, S., Gil-Garcia, J. R., Mellouli, S., Nahon, K., ... & Scholl, H. J. (2012). Understanding smart cities: An integrative framework. *In System Science (HICSS), 2012 45th Hawaii International Conference on* (pp. 2289-2297). IEEE.

Cioffi-Revilla, C. (2010). Computational social science. *Wiley Interdisciplinary Reviews: Computational Statistics*, 2(3), 259-271.

Corburn, J. (2003). Bringing local knowledge into environmental decision making improving urban planning for communities at risk. *Journal of Planning Education and Research*, 22(4), 420-433.

Coleman, D. J., Georgiadou, Y., & Labonte, J. (2009). Volunteered geographic information: The nature and motivation of produsers. *International Journal of Spatial Data Infrastructures Research*, 4(1), 332-358.

Craglia, M. (2007). Volunteered Geographic Information and Spatial Data Infrastructures: when do parallel lines converge. *Position paper for the VGI Specialist Meeting*, December 2007, Santa Barbara, USA (pp. 13-14).

Craglia, M. de Bie, K., Jackson, D., Pesaresi, M., Remetey-Fülöpp, G., Wang, C., Annoni, A. Bian, L.; Campbell, F, Ehlers, M., van Genderen, J., Goodchild, M. Guo, H., Lewis, A. Simpson, R., Skidmore, A., Woodgate, P. (2012). Digital Earth 2020: towards the vision for the next decade. *International Journal of Digital Earth*, 5(1), 4-21.

Craglia, M., & Annoni, A. (2007). INSPIRE: An innovative approach to the development of spatial data infrastructures in Europe. *Research and Theory in Advancing Spatial Data Infrastructure Concepts*, 93-105.

Craglia, M., Goodchild, M.F., Annoni, A., Camara, G., Gould, M., Kuhn, W., Mark, D., Masser, I., Maguire, D., Liang, S., Parsons, E. (2008). Next-Generation Digital Earth. *International Journal of Spatial Data Infrastructures Research*, 2008, (3),146-167.

Craglia, M., Ostermann, F., Spinsanti, L., 2012 B. Digital earth from vision to practice: making sense of citizen-generated content. *International Journal of Digital Earth,* 5(5), 398–416.

Cranshaw, J., Schwartz, R., Hong, J. I., & Sadeh, N. M. (2012). The Livehoods Project: Utilizing Social Media to Understand the Dynamics of a City. *In Sixth International AAAI Conference on Weblogs and Social Media ICWSM.*

Crompvoets, J., Bregt, A., Rajabifard, A., & Williamson, I. (2004). Assessing the worldwide developments of national spatial data clearinghouses*. International Journal of Geographical Information Science*, 18(7), 665-689.

Crooks, A., Croitoru, A., Stefanidis, A., & Radzikowski, J. (2013). # Earthquake: Twitter as a distributed sensor system. *Transactions in GIS*, 17(1), 124-147.

Dangermond, J. (2010). GeoDesign and GIS – Designing our Futures. In Buhmann, E. et al. (Eds.), *Peer Reviewed Proceedings of Digital Landscape Architecture.* Anhalt University of Applied Science, Germany.

Davidoff, P. (1965). Advocacy and pluralism in planning. *Journal of the american Institute of Planners*, 31(4), 331-338.

Davis Jr, C. A. (2008). Spatial Data Infrastructures. *Encyclopedia of Information Science and Technology*, 7, 3548-3553.

Davoudi, S. (2003). La partecipazione nella pianificazione per la sostenibilità [Participation in Planning for Sustainability]. *Urbanistica 2003*, 55, 119–129.

Dawes, S. S., & Pardo, T. A. (2002). Building collaborative digital government systems. *In Advances in digital government* (pp. 259-273). Springer US.

de Albuquerque, J. P., Herfort, B., Brenning, A., & Zipf, A. (2015). A geographic approach for combining social media and authoritative data towards identifying useful information for disaster management. *International Journal of Geographical Information Science*, 29(4), 667-689.

Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the royal statistical society*. *Series B (methodological)*, 1-38.

Dente, B. (1997). Una riflessione. *In Bonfiglioli, S., & Mareggi, M. (Eds.) Il tempo e la città fra natura e storia: atlante di progetti sui tempi della città*. Istituto Nazionale di Urbanistica. Urbanistica quaderni, n. 12, pp.178-180 (Roma, Inu Edizioni).

Directive 2007/2/EC, I. N. S. P. I. R. E. (2007). *Directive 2007/2/EC of the European Parliament and of the Council of 14 March 2007 establishing an Infrastructure for Spatial Information in the European Community (INSPIRE)*. Published in the official Journal on the 25th April.

Directive 2001/42/EC, S. E. A. (2001). *Directive 2001/42/EC of the European Parliament and of the Council of 27 June 2001 on the assessment of the effects of certain plans and programmes on the environment (OJ L 197, 21.7. 2001, pp. 30–37)*. Official Journal L, 197(21/07), 0030-0037.

Dirks, S., & Keeling, M. (2009). A Vision of Smarter Cities: How Cities Can Lead the Way into a Prosperous and Sustainable Future. *IBM Institute for Business Value*. Somers, NY. Retrieved from http://www-03.ibm.com/press/ attachments/IBV_Smarter_Cities_-_Final.pdf [Accessed 2014 March 26]

Dror, Y. (1986). Planning as fuzzy gambling: A radical perspective on coping with uncertainty. *Planning in turbulence*, 24-39.

Ellison, N. B. (2007). Social network sites: Definition, history, and scholarship. *Journal of Computer-Mediated Communication*, 13(1), 210-230.

Elwood, S. (2008 A). Volunteered geographic information: key questions, concepts and methods to guide emerging research and practice. *GeoJournal,* 72 (3–4), 133–135.

Elwood, S. (2010 B). Geographic information science: emerging research on the societal implications of the geospatial web. *Progress in Human Geography*, 34 (3), 349–357.

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: 10.1080/00045608.2011.595657.

Engler, N. J., Scassa, T., & Fraser Taylor, D. R. (2014). Cybercartography and volunteered geographic information. *Modern Cartography Series*, 5, 43-57.

Ervin, S. (2011). A system for GeoDesign. *Proceedings of Digital Landscape Architecture.* Anhalt University of Applied Science, 145-154.

ESRI (2015). *Tapestry™ Segmentation Reference Guide*. Retrieved from: http://www.esri.com/ library/whitepapers/pdfs/community-tapestry.pdf [Accessed 2015 September 11]

Ester, M., Kriegel, H. P., Sander, J., & Xu, X. (1996). A density-based algorithm for discovering clusters in large spatial databases with noise. In *Kdd* (Vol. 96, pp. 226-231).

Estivill-Castro, V. (2002). Why so many clustering algorithms: a position paper. *ACM SIGKDD explorations newsletter*, 4(1), 65-75. doi:10.1145/568574.568575.

Everitt, B. S., Landau, S. and Leese, M. (2001). *Cluster Analysis*. 4th edn. London: Arnold.

Executive Order 12906 (1994). *Coordinating Geographic Data Acquisition and Access: the National Spatial Data Infrastructure*. Retrieved from http://www.archives.gov/federal-register/executive-orders/pdf/12906.pdf [Accessed 2015 July 27]

Experian. (2010). *Optimise the value of your customers and locations, now and in the future. London: Experian*. Retrieved from: http://www.experian.co.uk/assets/business-strategies/brochures/mosaic-uk-2009-brochure-jun10.pdf [Accessed 2015 September 13]

FGDC (2004). *Geospatial one-stop: encouraging partnerships to enhance access to geospatial information*. Retrieved from: http://www.fgdc.gov/library/factsheets/documents/gos.pdf [Accessed 27 July 2015]

Fischer, F. (2000). *Citizens, experts, and the environment: The politics of local knowledge*. Duke University Press.

Fischer, F. (2012). VGI as Big Data: A new but delicate geographic data-source. *GeoInformatic*s, 15(3), 46-47.

Fischer, T. (2007). *Theory and Practice of Strategic Environmental Assessment*. London: Earthscan.

Fischer, T. B. (2010). Reviewing the quality of strategic environmental assessment reports for English spatial plan core strategies. *Environmental Impact Assessment Review*, 30(1), 62-69.

Flanagin, A. J., & Metzger, M. J. (2008). The credibility of volunteered geographic information. *GeoJournal*, 72(3-4), 137-148.

Flaxman, M. (2010). Fundamentals of Geodesign. *In: Buhmann, Pietsch, Kretzel, (Eds.). Peer Reviewed Proceedings Digital Landscape Architecture*. Anhalt University of Applied Science, Germany.

Floris, R., & Campagna, M. (2014). Social Media Geographic Information in Tourism Planning. *Tema. Journal of Land Use, Mobility and Environment*.

Foley, R., (2009). Integrated spatial data infrastructures. *In: Kitchin R. and Thrift N., (Eds.). The international encyclopaedia of human geography*. Vol. V. Elsevier, London, pp. 507–511.

Forgy, E. W. (1965). Cluster analysis of multivariate data: efficiency versus interpretability of classifications. *Biometrics*, 21, 768-769.

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)* (pp. 239-248). IEEE.

Friedmann J. (1973 A). *Retracking America: A Theory of Transactive Planning*. Garden City, NY, 1973. Doubleday/Anchor

Friedmann, J. (1987). *Planning in the public domain: From knowledge to action*. Princeton University Press. ISBN 0-691-07743-6.

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

GlobalWebIndex (2014). *GWI Social Q4 2014: the latest social networking trends*. Retrieved from: http://insight.globalwebindex.net/social-q1-2015. [Accessed 2015 July 7]

Göçmen, Z. A., & Ventura, S. J. (2010). Barriers to GIS use in planning. *Journal of the American Planning Association*, 76(2), 172-183.

Goldstein, H. A., & Carmin, J. (2006). Compact, diffuse, or would-be discipline? Assessing cohesion in planning scholarship, 1963-2002. *Journal of Planning Education and Research*, 26(1), 66-79.

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

Goodchild, M. F. (2010). Towards geodesign: Repurposing cartography and GIS?. *Cartographic Perspectives*, (66), 7-22.

Goodchild, M. F., & Li, L. (2012). Assuring the quality of volunteered geographic information. *Spatial statistics*, 1, 110-120.

Goodchild, M. F., & Glennon, J. A. (2010). Crowdsourcing geographic information for disaster response: a research frontier*. International Journal of Digital Earth*, 3(3), 231-241.

Gouveia, C., & Fonseca, A. (2008). New approaches to environmental monitoring: the use of ICT to explore volunteered geographic information. *GeoJournal*, 72(3-4), 185-197.

Graebner, D., Zanker M., Fliedl, G., Fuchs M. (2012). Classification of Customer Reviews based on Sentiment Analysis. *In 19th Conference on Information and Communication Technologies in Tourism (ENTER*), Springer, Helsingborg, Sweden, 2012.

Grekousis, G., & Thomas, H. (2012). Comparison of two fuzzy algorithms in geodemographic segmentation analysis: The Fuzzy C-Means and Gustafson–Kessel methods. *Applied Geography*, 34, 125-136.

Haklay, M., (2010 A). How good is volunteered geographical information? A comparative study of OpenStreetMap and ordnance survey datasets. *Environment and Planning B: Planning and Design* 37 (4), 682–703.

Haklay, M. (2013). Citizen Science and Volunteered Geographic Information – overview and typology of participation. *In Sui, D.Z., Elwood, S. and M.F. Goodchild (Eds.). Crowdsourcing Geographic Knowledge: Volunteered Geographic Information (VGI) in Theory and Practice*. Berlin: Springer. pp 105-122 DOI: 10.1007/978-94-007-4587-2_7

Haklay, M., Basiouka, S., Antoniou, V., & Ather, A. (2010). How many volunteers does it take to map an area well? The validity of Linus' law to volunteered geographic information. *The Cartographic Journal*, 47(4), 315-322.

Haklay, M., & Weber, P. (2008). Openstreetmap: User-generated street maps. *Pervasive Computing*, IEEE, 7(4), 12-18.

Hall, P. (1988). *Cities of tomorrow*. Blackwell Publishers.

Hall, R. E. (2000). The vision of a smart city. In Proceedings of the 2nd International Life Extension Technology Workshop, Paris, France. Retrieved from http://ntl.bts.gov/lib/14000/14800/14834/DE2001773961.pdf [Accessed 2014 Mar 28]

Halvey, M. J., & Keane, M. T. (2007). An assessment of tag presentation techniques. *In Proceedings of the 16th international conference on World Wide Web* (pp. 1313-1314). ACM.

Harris, P. M., & Ventura, S. J. (1995). The integration of geographic data with remotely sensed imagery to improve classification in an urban area. *Photogrammetric engineering and remote sensing*, 61(8), 993-998.

Harris, R., Sleight, P., & Webber, R. (2005). *Geodemographics, GIS and neighbourhood targeting* (Vol. 7). John Wiley and Sons.

Harrison, C., Eckman, B., Hamilton, R., Hartswick, P., Kalagnanam, J., Paraszczak, J., Williams, P. (2010). Foundations for Smarter Cities. *IBM Journal of Research and Development* , Vol.54, no.4, pp.1-16.

Hart, J. F. (1982). The highest form of the geographer's art. *Annals of the Association of American Geographers* 72(1), 1-29.

Hartigan, J. A., & Wong, M. A. (1979). Algorithm AS 136: A k-means clustering algorithm. *Applied statistics*, 100-108.

Healey, P. (1993). The communicative turn in planning theory and its implications for spatial strategy formation. *In Fischer F. and Forester J. (Eds). The Argumentative Turn in Policy Analysis and Planning*, Durham, NC, Duke University Press, pp. 223-253.

Healey, P., McDougall, G., & Thomas, M. J. (1982). Theoretical debates in planning: towards a coherent dialogue. *Planning Theory. Prospect for the 1980s*.

Hilbert, M. (2015). *e-Science for Digital Development: ICT4ICT4D*. Centre for Development Informatics, SEED, University of Manchester. Retrieved from: http://www.seed.manchester.ac.uk/medialibrary/IDPM/working_papers/di/di-wp60.pdf. [Accessed 2015 September 20]

Hudson-Smith, A., & Crooks, A. (2008). The renaissance of geographic information: neogeography, gaming and second life. *UCL Working Paper Series. Paper 142 – August 2008*. ISSN 1467-1298. Retrieved from: http://discovery.ucl.ac.uk/178942/1/PDF_142.pdf. [Accessed 2015 October 18]

Innes, J. E. (1995). Planning theory's emerging paradigm: communicative action and interactive practice. *Journal of planning education and research*, 14(3), 183-189.

Instagram Press (2015). *Stats*. Retrieved from: https://instagram.com/press/. [Accessed 2015 August 12].

Internet Live Stats (2015). *Google Search Statistics*. Retrieved from: http://www.internetlivestats.com/google-search-statistics/. [Accessed 2015 August 12].

ISO/TC 211, (2002). 19113—Geographic Information—Quality Principles. *International Organization for Standardization (ISO).* Geneva (CHE).

Jankowski, P., Andrienko, N., Andrienko, G., & Kisilevich, S. (2010). Discovering Landmark Preferences and Movement Patterns from Photo Postings. *Transaction in GIS*, 14(6), 833–852. doi:10.1111/j.1467-9671.2010.01235.x

Jennex, M. E. (2010). Implementing Social Media in Crisis Response Using Knowledge Management. *International Journal of Information Systems for Crisis Response and Management (IJISCRAM)*, 2(4), 20-32.

Jiang, M., & McGill, W. L. (2010). Human-centered sensing for crisis response and management analysis campaigns. *In Proceedings of Information Systems for Crisis Response and Management*, *ISCRM*, 10.

Johnson, S. C. (1967). Hierarchical clustering schemes. *Psychometrika*, 32(3), 241-254.

Kang, J. H., & Lerman, K. (2012). Using lists to measure homophily on twitter. In *AAAI Workshops*.

Kaser, O., & Lemire, D. (2007). Tag-cloud drawing: Algorithms for cloud visualization. In Proceedings of Tagging and Metadata for Social Information Organization (WWW 2007). http://arxiv.org/abs/cs/0703109

Kemp, S. (2015). Digital, Social & Mobile in 2015. We are social's compendium of global digital statistics. Retrieved from: http://www.slideshare.net/wearesocialsg/digital-social-mobile-in-2015?ref=http://wearesocial.it/. Accessed [2015 July 8]

Keßler, C., Trame, J., & Kauppinen, T. (2011). Tracking editing processes in volunteered geographic information: The case of OpenStreetMap. *In Proceedings of Workshop on Identifying Objects, Processes and Events in Spatio-Temporally Distributed Data (IOPE), Belfast (USA), 12 September 2011, 7 p.*

Khakee, A. (1998). Evaluation and planning: inseparable concepts. *Town Planning Review*, 69(4), 359.

Khakee, A., Barbanente, A., & Borri, D. (2000). Expert and experiential knowledge in planning. *Journal of the Operational Research Society*, 776-788.

Kietzmann, J. H., Hermkens, K., McCarthy, I. P., & Silvestre, B. S. (2011). Social media? Get serious! Understanding the functional building blocks of social media. *Business horizons*, 54(3), 241-251.

Klosterman, R. E. (1983). Fact and value in planning. *Journal of the American Planning Association*, 49(2), 216-225.

Knudsen, A. M. S., & Kahila, M. (2012). The role of Volunteered Geographic Information in participatory planning. Examples from Denmark and Finland. *Geoforum Perspektiv*, 11(21).

Kroes, N. (2010). The critical role of cities in making the Digital Agenda a reality. *Closing speech to Global Cities Dialogue Spring Summit of Mayors, European Commission-SPEECH/10/272*, Brussels, 28.

Krulwich, B. (1997). Lifestyle finder: Intelligent user profiling using large-scale demographic data. *AI magazine*, 18(2), 37.

Krumm, J., Davies, N. and Narayanaswami, C. (2008). User-generated content. *IEEE Pervasive Computing*, 7(4), 10-11.

Lamantia, J. (2007). *Text clouds: a new form of tag cloud*. Retrieved from: http://www.joelamantia. com/blog/archives/tag_clouds/text_clouds_a_new_fo rm_of_tag_cloud. html. [Accessed 2014 June 10].

Latonero, M., Shklovski, I. (2010). Respectfully yours in safety and service: emergency management and social media evangelism. *In Proceedings of the 7th International ISCRAM Conference*, Seattle, USA, May.

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*, Vol. 323, pp. 721–723.

Leslie, E., Coffee, N., Frank, L., Owen, N., Bauman, A., & Hugo, G. (2007). Walkability of local communities: using geographic information systems to objectively assess relevant environmental attributes. *Health & place*, 13(1), 111-122.

Li, R., Lei, K. H., Khadiwala, R., & Chang, K. C.-C. (2012 A). TEDAS: a Twitter Based Event Detection and Analysis System. *In IEEE 28th International Conference on Data Engineering (p. 3).*

Li, R., Wang, S., Deng, H., Wang, R., & Chang, K. C. C. (2012 B). Towards social user profiling: unified and discriminative influence model for inferring home locations. *In Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining* (pp. 1023-1031). ACM.

Lindblom, C.E. (1990). *Inquiry and Change: The Troubled Attempt to Understand and Shape Society*. Yale University Press: New Haven.

Lindskog, H. (2004). Smart communities initiatives. In Proceedings of the 3rd ISOneWorld Conference, pp. 16.

Liu, S. B., Palen, L., Sutton, J., Hughes, A. L., & Vieweg, S. (2008). In Search of the Bigger Picture: The Emergent Role of On-Line Photo Sharing in Times of Disaster. *In* F. Fiedrich and B. Van de Walle, (Eds.). *Proceeding of the Information Systems for Crisis Response and Management Conference (p. 10)*. Washington, DC, USA:

Liu, S., Palen, L., 2010. The new cartographers: crisis map mashups and the emergence of nongeographic practice. *Cartography and Geographic Information Science,* 37(1), 69–90.

Longley, P.A., Adnan, M., Lansley, G. (2014). Spatio-temporal demographic classification of the Twitter users. Extendend abstract. *In Eighth International Conference on Geographic Information Science*. Vienna, Austria, 23-26 September 2014.

Lorr, M. (1983). *Cluster Analysis for the Social Sciences*. San Francisco: Jossey–Bass.

Luce, L. (2012). Python, Twitter statistics and the 2012 French presidential election. In: Technical writeup. Retrieved from: http://www.laurentluce.com/posts/python-twitter-statistics-and-the-2012-french-presidential-election/ [Accessed 2013 Nov 14].

MacKay, D. (2003). An example inference task: clustering. *Information Theory, Inference and Learning Algorithms*, 284-292.

MacQueen, J. (1967). Some methods for classification and analysis of multivariate observations. *In Proceedings of the fifth Berkeley symposium on mathematical statistics and probability,* 1(14), 281-297.

Maguire, D. J., & Longley, P. A. (2005). The emergence of geoportals and their role in spatial data infrastructures. *Computers, environment and urban systems*, 29(1), 3-14.

Mandel, B., Culotta, A., Boulahanis, J., Stark, D., Lewis, B., & Rodrigue, J. (2012). A demographic analysis of online sentiment during hurricane Irene. *In Proceedings of the Second Workshop on Language in Social Media (pp. 27-36)*. Association for Computational Linguistics.

Manigas, L., Beneventi, M., & Vinelli, R. (2010). I dati geografici aperti ai cittadini, ai professionisti e alle Pubbliche Amministrazioni: il SITR-IDT della Regione Sardegna. *In Proceedings of the 14th National Conference ASITA*, Brescia November (pp. 9-12).

Manovich, L. (2010 A). Database as symbolic form. *Museums in a digital age*, 64-71.

Manovich, L. (2011 B). Trending: the promises and the challenges of big social data. *Debates in the digital humanities*, 460-475.

Mapping Science Committee. (1993). *Toward a coordinated spatial data infrastructure for the nation*. National Academies Press.

March, J. G. (1994). *Primer on decision making: How decisions happen*. Simon and Schuster.

Marcus, A., Bernstein, M. S., Badar, O., Karger, D. R., Madden, S., & Miller, R. C. (2011). TwitInfo: Aggregating and Visualizing Microblogs for Event Exploration. *In Proceedings of the 2011 annual conference on Human factors in computing systems (CHI '11)* (pp. 227– 236). New York, NY, USA: Association for Computing Machinery (ACM). doi:10.1145/1978942.1978975

Mareggi, M. (2002). Innovation in urban policy: the experience of Italian urban time policies. *Planning Theory & Practice*, 3(2), 173-194.

Mashhadi, A., Quattrone, G., Capra, L., & Mooney, P. (2012). On the accuracy of urban crowd-sourcing for maintaining large-scale geospatial databases. *In Proceedings of the Eighth Annual International Symposium on Wikis and Open Collaboration (p. 15)*. ACM.

Masser, I. (1999). All shapes and sizes: the first generation of national spatial data infrastructures. *International Journal of Geographical Information Science*, 13(1), 67-84.

Mathioudakis, M., & Koudas, N. (2010). TwitterMonitor: Trend Detection over the Twitter Stream. *In Proceeding of the2010 ACM SIGMOD International Conference on Management of Data (pp. 1155–1558*). Indianapolis, Indiana, USA.

Maué, P. (2007). Reputation as tool to ensure validity of VGI. *In Workshop on volunteered geographic information 2007*.

McDougall, K. (2009). Volunteered geographic information for building SDI. *In Proceedings of the Surveying and Spatial Sciences Institute Biennial International Conference (SSC 2009) (pp. 645-653)*. Surveying and Spatial Sciences Institute, Adelaide, Australia.

McPherson, M., Smith-Lovin, L., & Cook, J. M. (2001). Birds of a feather: Homophily in social networks. *Annual review of sociology*, 415-444.

Michie, D., Spiegelhalter, D. J., & Taylor, C. C. (1994). *Machine learning, neural and statistical classification*. Oxford: Oxford University Press.

Miguéns, J., Baggio, R., & Costa, C. (2008). Social media and tourism destinations: TripAdvisor case study. *Advances in Tourism Research*, 26-28.

Miller, C. (2006). A beast in the field: The Google Maps mashup as GIS. *Cartographica*, 41, 1878–1899.

Miller, H. J., & Han, J. (2009). *Geographic data mining and knowledge discovery*. CRC Press.

Milligan, G. W. (1996). Clustering validation: results and implications for applied analyses. *In Arabie, P., Hubert, L.J., and De Soete, G. (Eds.) Clustering and Classification*. Singapore: World Scientific Press.

Milligan, G. W., & Cooper, M. C. (1988). A study of standardization of variables in cluster analysis. *Journal of classification*, 5(2), 181-204.

Mooney, P., & Corcoran, P. (2012). Who are the contributors to OpenStreetMap and what do they do. *In Proceedings of 20th Annual GIS Research UK*, Lancaster, UK, 11-13.

Mooney, P., Corcoran, P., & Winstanley, A. (2010). A study of data representation of natural features in OpenStreetMap. *In Proceedings of GIScience (Vol. 150),* , Zurich (CHE), 14-17 September, pp. 150-156.

Nam, T., & Pardo, T. A. (2011). Smart city as urban innovation: Focusing on management, policy, and context. *In Proceedings of the 5th International Conference on Theory and Practice of Electronic Governance (pp. 185-194)*. ACM.

Napolitano, M., & Mooney, P. (2012). MVP OSM: a tool to identify areas of high quality contributor activity in OpenStreetMap. *The Bulletin of the Society of Cartographers*, 45(1), 10-18.

Nebert, D. (2004). *Developing Spatial Data Infrastructures: The SDI Cookbook*. Global Spatial Data Infrastructure. Open Geospatial Consortium.

Neis, P., Zielstra, D., & Zipf, A. (2011). The street network evolution of crowdsourced maps: OpenStreetMap in Germany 2007–2011. *Future Internet*, 4(1), 1-21.

Nogueras-Iso, J., Zarazaga-Soria, F.J., Lacasta, J., Béjar, R., & Muro-Medrano, P.R. (2004). Metadata standard interoperability: application in the geographic information domain. *Computers, environment and urban systems*, 28(6), 611-634.

Nonaka, I., Takeuchi, H. (1995). *The Knowledge-Creating Company. How Japanese Companies Create the Dynamics of Innovation*. Oxford University Press.

Noulas, A., Scellato, S., Mascolo, C., & Pontil, M. (2011). Exploiting Semantic Annotations for Clustering Geographic Areas and Users in Location-based Social Networks. *The Social Mobile Web*, 11.

Obermeyer, N. (2007). Thoughts on volunteered (geo) slavery. In Workshop on Volunteered Geographic Information, Santa Barbara, CA. Retrieved from http://www.ncgia.ucsb.edu/projects/vgi/docs/position/Obermeyer_Paper.pdf [Accessed 2015 August 5]

Oja, E., & Kaski, S. (1999). *Kohonen maps*. Elsevier.

Openshaw, S., & Gillard, A. (1978). On the stability of a spatial classification of census enumeration districts data. *In Batey P.W.S. (Eds.), Theory and methods in urban and regional analysis, (pp. 101-119)*. London: Pion.

O'Reilly, T. (2007). What is Web 2.0: Design patterns and business models for the next generation of software. *Communications & strategies*, (1), 17.

Ostermann, F. O., Huang, H., Andrienko, G., Andrienko, N., Capineri, C., Farkas, K., & Purves, R. S. (2015). Extracting and Comparing Places Using Geo-social Media. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Volume II-3/W5.

Paolillo, P. L., Benedetti, A., & Terlizzi, L. (2009). New survey instruments: studies for the environmental assessment report of the general plan in a case in Lombardy. *Planning, complexity and New ICT*, Alinea, Firenze, 215-224.

Parker, J. (2007). Strategic Environmental Assessment and SFs Operational Programmes: An assessment. *DG Environment European Commission Presentation at the Aarhus Workshop on Public Participation in Strategic Decision-Making (PPSD),* 03-04 December 2007, Sofia, Bulgaria.

Pearson, K. (1895). Note on regression and inheritance in the case of two parents. *Proceedings of the Royal Society of London*, 240-242.

Petersen, J., Gibin, M., Longley, P., Mateos, P., Atkinson, P., & Ashby, D. (2011). Geodemographics as a tool for targeting neighbourhoods in public health campaigns. *Journal of Geographical Systems*, 13(2), 173-192.

Phillips, A., Williamson, I., & Ezigbalike, C. (1999). Spatial data infrastructure concepts. *Australian Surveyor*, 44(1), 20-28.

Pohl, J., & Pohl, K. J. (2013). Big Data: Immediate Opportunities and Longer Term Challenges. *Proceedings of InterSymp-2013*, pp.12, Germany.

Poser, K., & Dransch, D. (2010). Volunteered geographic information for disaster management with application to rapid flood damage estimation. *Geomatica*, 64(1), 89-98.

Purves, R.S., Derungs, C. (2015). From space to place: place-based explorations of texts. *International Journal of Humanities and Arts Computing*, 9(1), pp. 74–94.

Rajabifard, A., M.-E. F. Feeney, and I. Williamson. (2003). Spatial data infrastructures: Concept, nature, and SDI hierarchy. *In Williamson, I., Rajabifard, A., Feeney M.E.F. (Eds.), Developing spatial data infrastructures: from concept to reality,* pp.17–40. London: Taylor & Francis.

Rambaldi, G., Kyem, P. A. K., McCall, M., & Weiner, D. (2006). Participatory spatial information management and communication in developing countries. *The Electronic Journal of Information Systems in Developing Countries*, 25.

Rantanen, H., & Kahila, M. (2009). The SoftGIS approach to local knowledge. *Journal of environmental management*, 90(6), 1981-1990.

Rattenbury, T., Good, N., & Naaman, M. (2007). Towards Automatic Extraction of Event and Place Semantics from Flickr Tags. *In Proceeding of the 30th International ACM SIGIR Conference on Research and Development in Information Retrieval (pp. 103–110).* New York, NY, USA.

Raymond, E.S. (1999). *The Cathedral and the Bazaar: Musings on Linux and Open Source by an Accidental Revolutionary*. O'Reilly, Beijing.

RDA Research (2010) geoSmart. Smarter geodemographic segments. Retrieved from: http://mailinglists. net.au/geoSmart_User_Guide.pdf [Accessed 2015 September 6].

Roche, S., Nabian, N., Kloeckl, K., & Ratti, C. (2012 A). Are 'smart cities' smart enough. *In Global geospatial conference* (pp. 215-235), Global Spatial Data Infrastructure Association.

Roche, S., Propeck-Zimmermann, E., & Mericskay, B. (2013 B). GeoWeb and crisis management: Issues and perspectives of volunteered geographic information. *GeoJournal*, 78(1), 21-40.

Roick, O. and Heuser, S. (2013). Location based social networks – definition, current state of the art and research Agenda. *Transactions in GIS*, 17 (5), 763–784.

Romesburg, C. (2004). *Cluster analysis for researchers*. Lulu.com.

Roth, E. (2013). Interactive mapping: what we know and what we need to know. *Journal of Spatial Information Science*, 6, 59–115.

Sakaki, T., Okazaki, M., & Matsuo, Y. (2010). Earthquake shakes Twitter users: real-time event detection by social sensors. *In Proceedings of the 19th international conference on World wide web (pp. 851-860*). ACM.

Scassa, T., Sattler, A. (2011). Location-based services and privacy. Canadian Journal of Law and Technology 9, (2), 99–134.

Scellato, S., Noulas, A., Lambiotte, R., & Mascolo, C. (2011). Socio-Spatial Properties of Online Location-Based Social Networks*. In Fifth International AAAI Conference on Weblogs and Social Media*.

Schaffers, H., Guzman, J. G., Navarro, M., & Merz, C. (2010). Living Labs for Rural Development. *Results from the C@ R Integrated Project, TRAGSA*, Madrid, Spain.

Scholl, H. J., Barzilai-Nahon, K., Ann, J. H., Popova, O. H., & Re, B. (2009). E-Commerce and e-Government: How do they Compare? what can they Learn from each Other?. *In 42nd Hawaii International Conference on System Sciences, 2009, HICSS'09, (pp. 1-10).* IEEE.

Sheate, W., Byron, H. and Smith, S. (2004). Implementing the SEA Directive: sectorial challenges and opportunities for the UK and EU. *European Environment*, 14, 73–93.

Sieber, R. (2007). Geoweb for social change. *Position paper in Workshop on Volunteered Geographic Information, December 13-14, 2007*. Santa Barbara, CA, USA. Retrieved from http://www.ncgia.ucsb.edu/projects/vgi/docs/supp_docs/Sieber_paper.pdf [Accessed 2015 August 10].

Silva, T. H., Melo, P. O., Almeida, J. M., Salles, J., & Loureiro, A. A. (2013 A). A picture of Instagram is worth more than a thousand words: Workload characterization and application. *In Proceedings of 2013 IEEE International Conference on Distributed Computing in Sensor Systems (DCOSS)*, (pp. 123-132). IEEE.

Silva, T. H., Vaz de Melo, P. O., Almeida, J. M., Salles, J., & Loureiro, A.A. (2013 B). A comparison of Foursquare and Instagram to the study of city dynamics and urban social behavior. *In Proceedings of the 2nd ACM SIGKDD International Workshop on Urban Computing* (p. 4). ACM.

Singleton, A. D., & Longley, P. A. (2009). Creating open source geodemographics: Refining a national classification of census output areas for applications in higher education. *Papers in Regional Science*, 88(3), 643-666.

Sinnappan, S., Farrell, C., & Stewart, E. (2010). Priceless tweets! A study on Twitter messages posted during crisis: Black Saturday. In Proceedings of the 21st Australasian Conference on Information Systems (ACIS).

Sleight, P. (1997). *Targeting customers: How to use geodemographic and lifestyle data in your business*. Henley-on-Thames: NTC Publications.

Spinsanti, L., & Ostermann, F. (2013). Automated geographic context analysis for volunteered information. *Applied Geography*, 43, 36-44.

Starbird, K., Palen, L., Hughes, A. L., & Vieweg, S. (2010). Chatter on the red: what hazards threat reveals about the social life of microblogged information. *In Proceedings of the 2010 ACM conference on Computer supported cooperative work (pp. 241-250)*. ACM. Savannah, Georgia, USA.

Stefanidis, A., Crooks, A., & Radzikowski, J. (2013). Harvesting ambient geospatial information from social media feeds. *GeoJournal*, 78(2), 319-338. http://dx.doi.org/10.1007/s10708-011-9438-2

Steinitz, C. (2012). *A Framework for Geodesign. Changing Geography by Design*. Esri Press, Redlands.

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

Sunstein, C. R. (2009). *Going to extremes: How like minds unite and divide*. Oxford University Press.

Tait, M. G. (2005). Implementing geoportals: applications of distributed GIS. *Computers, Environment and Urban Systems*, 29(1), 33-47.

Taylor, D.R.F. (2005). Cybercartography: Theory and Practice. First edition. Elsevier, Amsterdam.

Throgmorton, J. A. (1992). Planning as persuasive storytelling about the future: Negotiating an electric power rate settlement in Illinois. *Journal of Planning Education and Research*, 12(1), 17-31.

Thurm, S. & Kane, S.T.Y.I. (2010). Your Apps Are Watching You. *The Wall Street Journal.*

Tobler, W. R. (1970). A computer movie simulating urban growth in the Detroit region. *Economic geography*, 234-240.

Toppeta, D. (2010). The Smart City vision: How Innovation and ICT can build smart," liveable", sustainable cities. *The Innovation Knowledge Foundation. Think*. Retrieved from: http://www.intaaivn.org/images/cc/Urbanism/background %20documents/Toppeta_Report_005_2010.pdf [Accessed 2014 Mar 27].

Torres, Y.Q.A., Costa, L.M.A. (2014). Digital narratives: mapping contemporary use of urban open spaces through geo-social data. *In Proceeding of 12th International Conference on Design and Decision Support Systems in Architecture and Urban Planning, DDSS 2014*. 25-27 August 2014 Eindhoven, The Netherlands.

Tsoukas, H. (2006). *Complex Knowledge: Studies in Organizational Epistemology*. University Press, Oxford.

Tumasjan, A., Sprenger, T. O., Sandner, P. G., & Welpe, I. M. (2010). Predicting Elections with Twitter: What 140 Characters Reveal about Political Sentiment. *Proceedings of the Fourth International AAAI Conference on Weblogs and Social Media*, 10, 178-185.

Turner, A. (2006). *Introduction to Neogeography*. O'Reilly, Sebastopol, CA, USA.

Twitter Usage (2015). *Company facts*. Retrieved from: https://about.twitter.com/company. [Accessed 2015 August 12].

United Nations (UN) (1992). *Results of the World Conference on Environment and Development: Agenda 21*. UNCED, Rio de Janeiro, United Nations, New York.

United Nations General Assembly (UNGA) (1992). Rio Declaration on Environment and Development. *Report of the UN Conference on Environment and Development*, Rio de Janeiro.

Van Der Walt, C., & Barnard, E. (2006). Data characteristics that determine classifier performance. *SAIEE*, 98(3), 87-93. Retrieved from: http://spiderwebmail-home-backup.googlecode.com/svn/trunk/pesquisa/classifier_characteristics/cvdwalt_data_characteristics_SAIEE.pdf [Accessed 2015 October 20].

Van Exel, M., & Dias, E. (2011). Towards a methodology for trust stratification in VGI. *In Volunteered geographic information (VGI)—research progress and new developments*. Presented at the association of American geographers annual meeting. 3 April 2011, Seattle, USA.

Van Exel, M., Dias, E., & Fruijtier, S. (2010). The impact of crowdsourcing on spatial data quality indicators. In Proceedings of the Sixth International Conference on Geographic Information Science (GIScience 2010), 14-17 September, Zurich, CHE.

Vickers, D., & Rees, P. (2007). Creating the UK National Statistics 2001 output area classification. *Journal of the Royal Statistical Society: Series A (Statistics in Society), 170*(2), 379-403.

Vickers, D., Rees, P., & Birkin, M. (2005). Creating the national classification of census output areas: data, methods and results. Working Paper February 2015. School of Geography, University of Leeds, Leeds. Retrieved from http://www.geog.leeds.ac.uk/wpapers/05-2.pdf [Accessed 2015 July 8]

Vieweg, S., Hughes, A. L., Starbird, K., & Palen, L. (2010). Microblogging During Two Natural Hazards Events: What Twitter May Contribute to Situational Awareness. *In Proceedings of the 2010 annual conference on Human factors in computing systems (CHI '10) (pp. 1079–1088)*. Atlanta, GA, USA.

Voas, D., & Williamson, P. (2001). The diversity of diversity: a critique of geodemographic classification. *Area*, 33(1), 63-76.

Voronoi, G. (1909). Nouvelles applications des paramètres continus à théorie des formes quadratiques. Deuxième Mémoire. Recherches sur les paralléloèdres primitifs. *Journal für die reine und angewandte Mathematik*, 136, 67-182.

Washburn, D. & Sindhu, U. (2009). *Helping CIOs Understand "Smart City" Initiatives", Making Leaders Successful Every Day*. Forrester Research, Inc., Cambridge, MA, USA. Retrieved from http://public.dhe.ibm.com/partnerworld/pub/smb/smarterplanet/forr_help_cios_und_smart_city_initiatives.pdf. [Accessed 2014 Mar 21].

Wates, N. (2014). *The Community Planning Handbook: How people can shape their cities, towns & villages in any part of the world*. Routledge.

Webber, R. (2004). Targeting Customers: How to Use Geodemographic and Lifestyle Data in Your Business. *Interactive Marketing, 6*(2), 200-201.

Webber, R. J., & Craig, J. A. (1978). *Socio-economic classification of Local Authority Areas (Vol. 35)*. Stationery Office. London: Office of Population Censuses and Surveys.

Weiss, M. J. (2000). *The Clustered World*. New York: Little Brown.

Williams, S. (2007). *Application for GIS specialist meeting*. http://www.ncgia.ucsb.edu/projects/vgi/participants.html. Accessed [2013 February 2].

Williamson, I. P., Rajabifard, A., & Feeney, M.E.F. (2004). *Developing spatial data infrastructures: from concept to reality*. CRC Press.

Wilson, P. (1997). *Smart Communities Guidebook*. Governor of California, CA, USA.

Yin, J., & Carswell, J. D. (2011). Touch2Query enabled mobile devices: a case study using OpenStreetMap and iPhone. *In Tanaka, K., Fröhlich, P., Kim, K.S. (Eds.) Web and Wireless Geographical Information Systems (pp. 203-218).* Springer Berlin Heidelberg.

YouTube Press (2015). *Statistics*. Retrieved from https://www.youtube.com/yt/press/statistics.html. [Accessed 2015 August 12].

Zanon, B. (2014). Planners' Technical Expertise: Changing Paradigms and Practices in the Italian Experience. *Planning Practice and Research*, 29(1), 75-95.

Zielstra, D., Zipf, A. (2010). A comparative study of proprietary geodata and volunteered geographic information for Germany, *In Proceedings of the 13th AGILE International Conference on Geographic Information Science*. Guimarães (PRT), 4 Jun 2010, 15 p.

Zin, T. T., Pyke, T., Hiromitsu, H., & Takashi, T. (2013). Knowledge based Social Network Applications to Disaster Event Analysis. *In Proceedings of the International MultiConference 2013 of Engineers and Computer Scientists IMECS*, Vol. 1, p. 6. Hong Kong.

Zook, M., Graham, M., Shelton, T., & Gorman, S. (2010). Volunteered geographic information and crowdsourcing disaster relief: a case study of the Haitian earthquake. *World Medical & Health Policy*, 2(2), 7-33. DOI: 10.2202/1948-4682.1069

Zoppi, C. (2012). *Governance, pianificazione e valutazione strategica: sviluppo sostenibile e governance nella pianificazione urbanistica*. Gangemi Editore SPA.

LIST OF ABBREVIATIONS AND ACRONYMS

AGI	Ambient Geographic Information
A-GI	Authoritative-Geographic Information
API	Application Programming Interfaces
CIC	Cagliari, I Care!
CSS	Computational Social Science
DB-SCAN	Density-Based Spatial Clustering of Application with Noise
DI	Digital Information
EM	Expectation-Maximization
FB-DBSCAN	Feature-Based Density-Based Spatial Clustering of Application with Noise
FGDC	Federal Geographic Data Committee
GI	Geographic Information
GIS	Geographic Information System
GOS	Geospatial One Stop
GPS	Global Positioning System
ICT	Information and Communication Technologies
ISTAT	Istituto Nazionale di Statistica
iVGI	inVoluntary Geographic Information
LBSN	Location-Based Social Networks
NSDI	National Spatial Data Infrastructure
OAC	Output Area Classification
OSM	Open Street Map
PIC	Place, I Care!
POI	Points Of Interest
PPGIS	Public Participation Geographic Information System

SA	Situational Awareness
SDI	Spatial Data Infrastructures
SEA	Strategic Environmental Assessment
SITR-IDT	Sistema Informativo Territoriale Regionale – Infrastrutura di Dati Territoriali
SMGI	Social Media Geographic Information
SNS	Social Network Sites
SOA	Service-Oriented Architecture
SOM	Self-Organizing Maps
SRSDI	Sardinia Region Spatial Data Infrastructure
STTx	Spatial-Temporal-Textual Analysis
SVM	Support Vector Machine
UGC	User-Generated Contents
UGSC	User-Generated Spatial Contents
VGI	Volunteered Geographic Information

LIST OF FIGURES AND TABLES

FIGURES

TABLES

ANNEXES

ANNEX 1

CORRELATIONS BETWEEN GEODEMOGRAPHIC CLASSIFICATION VARIABLES

A1.1 Description

The geodemographic classification developed in the thesis is based upon 43 socio-economic variables, generated from the official 199 variables provided by the Italian National Institute of Statistics (ISTAT) for the 2001 census datA1. After the variables' selection step, the resulting dataset is assessed in order to evaluate the potential presence of variables' pairs exposing strong correlations by means of the Pearson coefficient. This attachment provides the resulting correlation tables for each variable.

Variable 1: Percentage of resident population aged 0-4 years	Pearson Correlation	p-value
Var_V2	0.09918	0.0
Var_V3	0.16071	0.0
Var_V4	-0.27805	0.0
Var_V5	-0.20891	0.0
Var_V6	0.03938	0.00012
Var_V7	0.0302	0.00326
Var_V8	0.03139	0.00223
Var_V9	-0.03724	0.00029
Var_V10	-0.02078	0.04292
Var_V11	0.10457	0.0
Var_V12	-0.12859	0.0
Var_V13	-0.18523	0.0
Var_V14	-0.11495	0.0
Var_V15	0.25332	0.0
Var_V16	0.06561	0.0
Var_V17	-0.03121	0.00237
Var_V18	0.05542	0.0
Var_V19	0.07832	0.0
Var_V20	-0.02498	0.01495
Var_V21	0.15939	0.0
Var_V22	0.04899	0.0
Var_V23	0.00729	0.47749
Var_V24	-0.09953	0.0
Var_V25	0.11542	0.0
Var_V26	0.05102	0.0
Var_V27	-0.03354	0.00108
Var_V28	-0.02714	0.00821
Var_V29	-0.02362	0.02143
Var_V30	0.11482	0.0
Var_V31	-0.09056	0.0
Var_V32	0.10192	0.0
Var_V33	0.05286	0.0
Var_V34	0.06586	0.0
Var_V35	-0.1233	0.0
Var_V36	-0.22809	0.0
Var_V37	0.0299	0.00358
Var_V38	0.03536	0.00057
Var_V39	0.05849	0.0

Var_V40	0.02209	0.03142
Var_V41	-0.01788	0.08165
Var_V42	0.02535	0.01355
Var_V43	-0.01982	0.05353

Variable 2: Percentage of resident population aged 5-14 years	Pearson Correlation	p-value
Var_V3	0.01351	0.18828
Var_V4	-0.26227	0.0
Var_V5	-0.30674	0.0
Var_V6	-0.02282	0.02621
Var_V7	-0.01561	0.12836
Var_V8	0.01798	0.0799
Var_V9	-0.05616	0.0
Var_V10	0.05536	0.0
Var_V11	0.07495	0.0
Var_V12	-0.19556	0.0
Var_V13	-0.31323	0.0
Var_V14	-0.18579	0.0
Var_V15	0.31036	0.0
Var_V16	0.28059	0.0
Var_V17	-0.02077	0.04306
Var_V18	0.05393	0.0
Var_V19	0.11376	0.0
Var_V20	0.03685	0.00033
Var_V21	0.23459	0.0
Var_V22	0.03	0.00347
Var_V23	0.07481	0.0
Var_V24	-0.15586	0.0
Var_V25	0.17164	0.0
Var_V26	0.07862	0.0
Var_V27	-0.06187	0.0
Var_V28	-0.07323	0.0
Var_V29	-0.13963	0.0
Var_V30	-0.05096	0.0
Var_V31	0.12745	0.0
Var_V32	0.29578	0.0
Var_V33	0.14227	0.0
Var_V34	0.1029	0.0
Var_V35	-0.01222	0.2338
Var_V36	-0.32832	0.0
Var_V37	-0.0085	0.40796
Var_V38	0.10918	0.0
Var_V39	0.01456	0.15614
Var_V40	0.07899	0.0
Var_V41	0.04435	2e-05
Var_V42	0.0203	0.04806
Var_V43	-0.05198	0.0

Variable 3: Percentage of resident population aged 25-44 years	Pearson Correlation	p-value
Var_V4	-0.55273	0.0
Var_V5	-0.37127	0.0
Var_V6	0.08285	0.0
Var_V7	0.04301	3e-05
Var_V8	0.0447	1e-05
Var_V9	-0.01396	0.17386
Var_V10	0.14579	0.0
Var_V11	-0.0451	1e-05

Var_V12	-0.14588	0.0
Var_V13	0.08491	0.0
Var_V14	-0.06634	0.0
Var_V15	0.01691	0.09951
Var_V16	-0.07878	0.0
Var_V17	-0.15865	0.0
Var_V18	0.14659	0.0
Var_V19	-0.02813	0.00615
Var_V20	-0.16865	0.0
Var_V21	0.01969	0.0551
Var_V22	-0.01083	0.29138
Var_V23	-0.11584	0.0
Var_V24	-0.10943	0.0
Var_V25	0.10256	0.0
Var_V26	0.02295	0.02536
Var_V27	-0.03148	0.00217
Var_V28	-0.00746	0.46771
Var_V29	-0.01312	0.20128
Var_V30	0.31307	0.0
Var_V31	-0.27673	0.0
Var_V32	0.14875	0.0
Var_V33	0.0374	0.00027
Var_V34	0.12925	0.0
Var_V35	-0.13947	0.0
Var_V36	-0.33167	0.0
Var_V37	0.08199	0.0
Var_V38	0.0176	0.08642
Var_V39	0.09415	0.0
Var_V40	0.0112	0.27517
Var_V41	-0.0371	0.0003
Var_V42	0.03581	0.00049
Var_V43	0.02125	0.03844

Variable 4: Percentage of resident population aged 45-64 years	Pearson Correlation	p-value
Var_V5	-0.21293	0.0
Var_V6	-0.03216	0.00173
Var_V7	-0.02233	0.0296
Var_V8	-0.01138	0.26781
Var_V9	-0.03845	0.00018
Var_V10	-0.07835	0.0
Var_V11	0.01861	0.06987
Var_V12	0.08697	0.0
Var_V13	0.1178	0.0
Var_V14	0.03819	0.0002
Var_V15	-0.10736	0.0
Var_V16	-0.09333	0.0
Var_V17	0.09521	0.0
Var_V18	-0.11416	0.0
Var_V19	-0.00824	0.42197
Var_V20	0.01335	0.19342
Var_V21	-0.12573	0.0
Var_V22	-0.01773	0.08418
Var_V23	0.02825	0.00593
Var_V24	-0.06939	0.0
Var_V25	0.06026	0.0
Var_V26	-0.0104	0.31091
Var_V27	-3e-05	0.99738
Var_V28	0.00404	0.694
Var_V29	0.08691	0.0

Var_V30	-0.04285	3e-05
Var_V31	-0.01161	0.2583
Var_V32	-0.06563	0.0
Var_V33	-0.04506	1e-05
Var_V34	-0.02372	0.02088
Var_V35	0.05431	0.0
Var_V36	0.08247	0.0
Var_V37	0.01506	0.14239
Var_V38	-0.06723	0.0
Var_V39	-0.03166	0.00204
Var_V40	-0.03737	0.00027
Var_V41	0.00371	0.71751
Var_V42	-0.01439	0.16102
Var_V43	0.01081	0.29255

Variable 5: Percentage of resident population aged > 65 years	Pearson Correlation	p-value
Var_V6	-0.03908	0.00014
Var_V7	-0.00792	0.44037
Var_V8	-0.05034	0.0
Var_V9	0.09722	0.0
Var_V10	-0.15268	0.0
Var_V11	-0.05282	0.0
Var_V12	0.30508	0.0
Var_V13	0.17918	0.0
Var_V14	0.25369	0.0
Var_V15	-0.28314	0.0
Var_V16	-0.1853	0.0
Var_V17	0.08922	0.0
Var_V18	-0.10398	0.0
Var_V19	-0.11379	0.0
Var_V20	0.0727	0.0
Var_V21	-0.15597	0.0
Var_V22	-0.01353	0.18764
Var_V23	-0.03156	0.00211
Var_V24	0.37543	0.0
Var_V25	-0.38385	0.0
Var_V26	-0.10493	0.0
Var_V27	0.09732	0.0
Var_V28	0.08273	0.0
Var_V29	0.09654	0.0
Var_V30	-0.42523	0.0
Var_V31	0.3301	0.0
Var_V32	-0.29714	0.0
Var_V33	-0.17824	0.0
Var_V34	-0.17843	0.0
Var_V35	-0.17458	0.0
Var_V36	0.61877	0.0
Var_V37	-0.04649	1e-05
Var_V38	-0.13688	0.0
Var_V39	-0.07608	0.0
Var_V40	-0.10347	0.0
Var_V41	0.01021	0.31991
Var_V42	-0.07569	0.0
Var_V43	0.0292	0.00444

Variable 6: Percentage of resident population Afrikans	Pearson Correlation	p-value
Var_V7	0.05513	0.0

Var_V8	-0.0027	0.79244
Var_V9	0.02292	0.02558
Var_V10	-0.00317	0.75753
Var_V11	0.0146	0.15495
Var_V12	-0.01756	0.08713
Var_V13	0.05483	0.0
Var_V14	0.00771	0.45277
Var_V15	-0.04704	0.0
Var_V16	-0.02466	0.0163
Var_V17	-0.10778	0.0
Var_V18	0.11269	0.0
Var_V19	-0.03732	0.00028
Var_V20	-0.0806	0.0
Var_V21	0.02815	0.00611
Var_V22	0.00103	0.91987
Var_V23	-0.06882	0.0
Var_V24	0.01139	0.26737
Var_V25	-0.01901	0.06402
Var_V26	0.02312	0.02434
Var_V27	-0.03254	0.00152
Var_V28	-0.00816	0.4266
Var_V29	-0.03199	0.00183
Var_V30	-0.00355	0.72933
Var_V31	0.02183	0.03351
Var_V32	0.00215	0.83413
Var_V33	-0.03889	0.00015
Var_V34	0.01214	0.23703
Var_V35	-0.04259	3e-05
Var_V36	-0.01872	0.06818
Var_V37	0.03508	0.00063
Var_V38	-0.0336	0.00106
Var_V39	0.01362	0.18457
Var_V40	-0.00971	0.3444
Var_V41	-0.05573	0.0
Var_V42	0.0649	0.0
Var_V43	-0.02876	0.00509

Variable 7: Percentage of resident population Asian	Pearson Correlation	p-value
Var_V8	-0.00512	0.61808
Var_V9	0.04507	1e-05
Var_V10	0.02465	0.01633
Var_V11	-0.02795	0.00648
Var_V12	0.00626	0.54191
Var_V13	0.04009	9e-05
Var_V14	-0.00795	0.4387
Var_V15	-0.04136	6e-05
Var_V16	-0.0172	0.0939
Var_V17	-0.06719	0.0
Var_V18	0.05693	0.0
Var_V19	-0.06881	0.0
Var_V20	-0.02771	0.00695
Var_V21	-0.00457	0.65621
Var_V22	0.00123	0.90489
Var_V23	-0.00901	0.38039
Var_V24	0.04983	0.0
Var_V25	-0.05791	0.0
Var_V26	-0.03925	0.00013
Var_V27	0.03065	0.00283
Var_V28	0.03656	0.00037

Var_V29	0.02176	0.03401
Var_V30	0.02381	0.0204
Var_V31	-0.03427	0.00084
Var_V32	0.02204	0.03179
Var_V33	-0.03733	0.00028
Var_V34	0.01027	0.31738
Var_V35	-0.02068	0.04396
Var_V36	-0.0064	0.53293
Var_V37	0.00413	0.68751
Var_V38	-0.00505	0.6227
Var_V39	-0.03001	0.00346
Var_V40	-0.01338	0.19244
Var_V41	-0.00165	0.87197
Var_V42	0.03064	0.00284
Var_V43	0.0248	0.01571

Variable 8: Percentage of resident population Caucasic of Hispanic	Pearson Correlation	p-value
Var_V9	-0.05857	0.0
Var_V10	0.00456	0.65709
Var_V11	0.01755	0.0874
Var_V12	-0.03342	0.00113
Var_V13	0.04586	1e-05
Var_V14	0.04115	6e-05
Var_V15	-0.07986	0.0
Var_V16	0.00664	0.51808
Var_V17	-0.06846	0.0
Var_V18	0.05873	0.0
Var_V19	0.00037	0.97141
Var_V20	-0.03637	0.0004
Var_V21	0.12016	0.0
Var_V22	0.19425	0.0
Var_V23	-0.04421	2e-05
Var_V24	-0.02713	0.00821
Var_V25	0.02794	0.0065
Var_V26	0.05868	0.0
Var_V27	-0.05901	0.0
Var_V28	-0.03453	0.00077
Var_V29	0.02709	0.00833
Var_V30	0.05668	0.0
Var_V31	-0.06722	0.0
Var_V32	-0.04532	1e-05
Var_V33	-0.04165	5e-05
Var_V34	-0.00159	0.87725
Var_V35	-0.05595	0.0
Var_V36	-0.00429	0.67582
Var_V37	0.01747	0.08881
Var_V38	-0.05315	0.0
Var_V39	0.05415	0.0
Var_V40	-0.02158	0.03551
Var_V41	-0.06309	0.0
Var_V42	-0.00284	0.78212
Var_V43	-0.03794	0.00022

Variable 9: Density Pop/sqKm [normalized by range method]	Pearson Correlation	p-value
Var_V10	-0.04602	1e-05
Var_V11	-0.02692	0.00874

Var_V12	0.1087	0.0
Var_V13	-0.01307	0.20291
Var_V14	0.07179	0.0
Var_V15	0.00315	0.75902
Var_V16	-0.05425	0.0
Var_V17	-0.09195	0.0
Var_V18	0.09509	0.0
Var_V19	-0.32542	0.0
Var_V20	-0.03635	0.0004
Var_V21	-0.02357	0.02167
Var_V22	-0.01229	0.23145
Var_V23	-0.13471	0.0
Var_V24	0.32608	0.0
Var_V25	-0.32399	0.0
Var_V26	-0.57821	0.0
Var_V27	0.59323	0.0
Var_V28	0.57674	0.0
Var_V29	0.15782	0.0
Var_V30	-0.00156	0.87911
Var_V31	-0.09033	0.0
Var_V32	0.1434	0.0
Var_V33	-0.11956	0.0
Var_V34	0.0095	0.35506
Var_V35	0.1003	0.0
Var_V36	0.03322	0.00121
Var_V37	-0.19296	0.0
Var_V38	0.2633	0.0
Var_V39	-0.26431	0.0
Var_V40	-0.05088	0.0
Var_V41	0.2665	0.0
Var_V42	0.0789	0.0
Var_V43	0.18001	0.0

Variable 10: Percentage of unmarried	**Pearson Correlation**	**p-value**
Var_V11	-0.77737	0.0
Var_V12	-0.28825	0.0
Var_V13	0.23051	0.0
Var_V14	-0.34446	0.0
Var_V15	-0.12486	0.0
Var_V16	0.15495	0.0
Var_V17	-0.1211	0.0
Var_V18	0.06965	0.0
Var_V19	-0.07907	0.0
Var_V20	-0.13342	0.0
Var_V21	0.07298	0.0
Var_V22	0.03332	0.00117
Var_V23	-0.11636	0.0
Var_V24	-0.02116	0.03928
Var_V25	-0.00754	0.4626
Var_V26	0.00681	0.50721
Var_V27	-0.04051	8e-05
Var_V28	-0.07693	0.0
Var_V29	-0.01308	0.20259
Var_V30	-0.01125	0.27319
Var_V31	0.01782	0.08256
Var_V32	0.08178	0.0
Var_V33	-0.04311	3e-05
Var_V34	-0.00337	0.743
Var_V35	0.02622	0.01066

Var_V36	-0.09017	0.0
Var_V37	0.10132	0.0
Var_V38	-0.11116	0.0
Var_V39	0.21836	0.0
Var_V40	-0.06892	0.0
Var_V41	-0.06312	0.0
Var_V42	-0.1006	0.0
Var_V43	-0.07801	0.0

Variable 11: Percentage of married	Pearson Correlation	p-value
Var_V12	-0.37826	0.0
Var_V13	-0.40752	0.0
Var_V14	0.29394	0.0
Var_V15	0.28109	0.0
Var_V16	-0.05138	0.0
Var_V17	0.14262	0.0
Var_V18	-0.08592	0.0
Var_V19	0.17313	0.0
Var_V20	0.16024	0.0
Var_V21	0.01924	0.06093
Var_V22	-0.02517	0.0142
Var_V23	0.1736	0.0
Var_V24	-0.13342	0.0
Var_V25	0.16748	0.0
Var_V26	0.06907	0.0
Var_V27	-0.03074	0.00275
Var_V28	-0.01631	0.11205
Var_V29	-0.03079	0.0027
Var_V30	0.06102	0.0
Var_V31	-0.03753	0.00026
Var_V32	0.00658	0.52166
Var_V33	0.10705	0.0
Var_V34	0.0781	0.0
Var_V35	0.03465	0.00074
Var_V36	-0.08313	0.0
Var_V37	-0.04691	0.0
Var_V38	0.12864	0.0
Var_V39	-0.13586	0.0
Var_V40	0.11176	0.0
Var_V41	0.0339	0.00096
Var_V42	0.10332	0.0
Var_V43	0.04483	1e-05

Variable 12: Percentage of separated, divorced or widowed	Pearson Correlation	p-value
Var_V13	0.28111	0.0
Var_V14	0.05946	0.0
Var_V15	-0.24413	0.0
Var_V16	-0.14981	0.0
Var_V17	-0.03888	0.00015
Var_V18	0.0283	0.00584
Var_V19	-0.14719	0.0
Var_V20	-0.04759	0.0
Var_V21	-0.13669	0.0
Var_V22	-0.01072	0.29663
Var_V23	-0.09303	0.0
Var_V24	0.23424	0.0
Var_V25	-0.24385	0.0

Var_V26	-0.11516	0.0
Var_V27	0.1064	0.0
Var_V28	0.13804	0.0
Var_V29	0.06613	0.0
Var_V30	-0.07634	0.0
Var_V31	0.03089	0.00262
Var_V32	-0.13036	0.0
Var_V33	-0.09951	0.0
Var_V34	-0.11394	0.0
Var_V35	-0.09132	0.0
Var_V36	0.25924	0.0
Var_V37	-0.0777	0.0
Var_V38	-0.03224	0.00168
Var_V39	-0.11452	0.0
Var_V40	-0.06871	0.0
Var_V41	0.04128	6e-05
Var_V42	-0.00924	0.36824
Var_V43	0.04657	1e-05

Variable 13: Percentage of single person household	Pearson Correlation	p-value
Var_V14	-0.28556	0.0
Var_V15	-0.62609	0.0
Var_V16	-0.30628	0.0
Var_V17	-0.0991	0.0
Var_V18	0.07312	0.0
Var_V19	-0.17419	0.0
Var_V20	-0.31976	0.0
Var_V21	-0.28972	0.0
Var_V22	-0.00719	0.48382
Var_V23	-0.32684	0.0
Var_V24	0.13193	0.0
Var_V25	-0.14725	0.0
Var_V26	-0.00016	0.9873
Var_V27	-0.01776	0.08361
Var_V28	0.03091	0.0026
Var_V29	0.09073	0.0
Var_V30	-0.06471	0.0
Var_V31	0.00603	0.55693
Var_V32	-0.10202	0.0
Var_V33	-0.14666	0.0
Var_V34	-0.06088	0.0
Var_V35	-0.20686	0.0
Var_V36	0.19538	0.0
Var_V37	0.12728	0.0
Var_V38	-0.26122	0.0
Var_V39	0.12138	0.0
Var_V40	-0.16211	0.0
Var_V41	-0.08512	0.0
Var_V42	-0.0903	0.0
Var_V43	-0.00773	0.45149

Variable 14: Percentage of couple household	Pearson Correlation	p-value
Var_V15	-0.31204	0.0
Var_V16	-0.17548	0.0
Var_V17	0.08252	0.0
Var_V18	-0.04756	0.0
Var_V19	0.0102	0.32045
Var_V20	0.08219	0.0

Var_V21	-0.0917	0.0
Var_V22	-0.01253	0.22238
Var_V23	0.03643	0.00039
Var_V24	0.12654	0.0
Var_V25	-0.11002	0.0
Var_V26	-0.06641	0.0
Var_V27	0.08715	0.0
Var_V28	0.10124	0.0
Var_V29	0.06287	0.0
Var_V30	-0.11591	0.0
Var_V31	0.06873	0.0
Var_V32	-0.09359	0.0
Var_V33	-0.04259	3e-05
Var_V34	-0.03864	0.00017
Var_V35	-0.07453	0.0
Var_V36	0.16886	0.0
Var_V37	-0.03132	0.00228
Var_V38	0.00653	0.52492
Var_V39	-0.08389	0.0
Var_V40	-0.00625	0.54301
Var_V41	0.01468	0.15291
Var_V42	0.02247	0.02864
Var_V43	0.06814	0.0

Variable 15: Percentage of household of 3-4 persons	Pearson Correlation	p-value
Var_V16	-0.13826	0.0
Var_V17	0.0887	0.0
Var_V18	-0.02983	0.00366
Var_V19	0.19403	0.0
Var_V20	0.25955	0.0
Var_V21	0.22026	0.0
Var_V22	-0.01226	0.23253
Var_V23	0.27962	0.0
Var_V24	-0.15292	0.0
Var_V25	0.18546	0.0
Var_V26	0.03096	0.00256
Var_V27	0.00643	0.53129
Var_V28	-0.02747	0.00745
Var_V29	-0.08872	0.0
Var_V30	0.14718	0.0
Var_V31	-0.08211	0.0
Var_V32	0.1904	0.0
Var_V33	0.13483	0.0
Var_V34	0.12189	0.0
Var_V35	0.19485	0.0
Var_V36	-0.26465	0.0
Var_V37	-0.10202	0.0
Var_V38	0.23811	0.0
Var_V39	-0.09889	0.0
Var_V40	0.12967	0.0
Var_V41	0.0856	0.0
Var_V42	0.08532	0.0
Var_V43	0.01558	0.12922

Variable 16: Percentage of household > 5 persons	Pearson Correlation	p-value
Var_V17	0.01003	0.32871
Var_V18	0.0113	0.27102

Var_V19	0.10485	0.0
Var_V20	0.15119	0.0
Var_V21	0.3094	0.0
Var_V22	0.04895	0.0
Var_V23	0.16174	0.0
Var_V24	-0.10533	0.0
Var_V25	0.11381	0.0
Var_V26	0.07358	0.0
Var_V27	-0.06525	0.0
Var_V28	-0.11223	0.0
Var_V29	-0.1298	0.0
Var_V30	0.02122	0.03871
Var_V31	0.05618	0.0
Var_V32	0.02708	0.00834
Var_V33	0.10924	0.0
Var_V34	-0.0324	0.0016
Var_V35	0.16046	0.0
Var_V36	-0.15211	0.0
Var_V37	-0.00203	0.84294
Var_V38	0.07255	0.0
Var_V39	0.06414	0.0
Var_V40	0.10327	0.0
Var_V41	-0.03402	0.00092
Var_V42	0.01021	0.31979
Var_V43	-0.07941	0.0

Variable 17: Percentage of house owners	**Pearson Correlation**	**p-value**
Var_V18	-0.93341	0.0
Var_V19	0.29385	0.0
Var_V20	0.35701	0.0
Var_V21	-0.14419	0.0
Var_V22	0.00997	0.3315
Var_V23	0.35685	0.0
Var_V24	-0.09669	0.0
Var_V25	0.16398	0.0
Var_V26	0.1211	0.0
Var_V27	-0.04507	1e-05
Var_V28	-0.13461	0.0
Var_V29	0.03152	0.00213
Var_V30	-0.04004	0.0001
Var_V31	0.01805	0.0788
Var_V32	-0.05315	0.0
Var_V33	0.14273	0.0
Var_V34	0.11526	0.0
Var_V35	0.08838	0.0
Var_V36	0.06523	0.0
Var_V37	0.13371	0.0
Var_V38	-0.0714	0.0
Var_V39	0.04518	1e-05
Var_V40	0.02375	0.02072
Var_V41	-0.0263	0.01041
Var_V42	-0.01019	0.32119
Var_V43	0.05145	0.0

Variable 18: Percentage of house tenants	**Pearson Correlation**	**p-value**
Var_V19	-0.14437	0.0
Var_V20	-0.21978	0.0

Var_V21	0.1744	0.0
Var_V22	-0.00952	0.35373
Var_V23	-0.24805	0.0
Var_V24	0.12846	0.0
Var_V25	-0.13094	0.0
Var_V26	-0.07183	0.0
Var_V27	0.07048	0.0
Var_V28	0.16016	0.0
Var_V29	-0.04537	1e-05
Var_V30	0.04014	9e-05
Var_V31	-0.01009	0.32576
Var_V32	0.10002	0.0
Var_V33	-0.12138	0.0
Var_V34	-0.08174	0.0
Var_V35	-0.05346	0.0
Var_V36	-0.11604	0.0
Var_V37	-0.12815	0.0
Var_V38	0.10801	0.0
Var_V39	-0.0441	2e-05
Var_V40	0.00139	0.89267
Var_V41	0.02494	0.01513
Var_V42	0.03594	0.00046
Var_V43	-0.03688	0.00033

Variable 19: Percentage of house with heating	Pearson Correlation	p-value
Var_V20	0.37995	0.0
Var_V21	-0.02052	0.04563
Var_V22	-0.04738	0.0
Var_V23	0.36982	0.0
Var_V24	-0.21119	0.0
Var_V25	0.29185	0.0
Var_V26	0.28432	0.0
Var_V27	-0.1962	0.0
Var_V28	-0.25277	0.0
Var_V29	-0.05839	0.0
Var_V30	0.02436	0.01765
Var_V31	0.01181	0.24985
Var_V32	0.06866	0.0
Var_V33	0.12118	0.0
Var_V34	0.1043	0.0
Var_V35	0.08328	0.0
Var_V36	-0.11991	0.0
Var_V37	0.0863	0.0
Var_V38	-0.0257	0.0123
Var_V39	0.13416	0.0
Var_V40	0.08178	0.0
Var_V41	-0.09621	0.0
Var_V42	-0.03037	0.00309
Var_V43	-0.05064	0.0

Variable 20: Average rooms per house [normalized by range method]	Pearson Correlation	p-value
Var_V21	-0.22056	0.0
Var_V22	-0.02334	0.02303
Var_V23	0.80762	0.0
Var_V24	0.02739	0.00762
Var_V25	0.04683	1e-05

Var_V26	-0.0354	0.00056
Var_V27	0.12294	0.0
Var_V28	-0.01924	0.06098
Var_V29	0.2137	0.0
Var_V30	0.02174	0.03419
Var_V31	-0.14398	0.0
Var_V32	0.07253	0.0
Var_V33	0.09059	0.0
Var_V34	0.14851	0.0
Var_V35	0.26086	0.0
Var_V36	-0.02515	0.01429
Var_V37	0.03641	0.00039
Var_V38	0.07781	0.0
Var_V39	-0.07504	0.0
Var_V40	-0.009	0.38051
Var_V41	0.13645	0.0
Var_V42	0.0077	0.45352
Var_V43	0.15164	0.0

Variable 21: Average persons per room [normalized by range method]	Pearson Correlation	p-value
Var_V22	0.18902	0.0
Var_V23	-0.12602	0.0
Var_V24	-0.07072	0.0
Var_V25	0.07324	0.0
Var_V26	0.07725	0.0
Var_V27	-0.07598	0.0
Var_V28	-0.04174	5e-05
Var_V29	-0.16993	0.0
Var_V30	0.02318	0.02394
Var_V31	0.07772	0.0
Var_V32	0.04097	7e-05
Var_V33	0.02996	0.00352
Var_V34	-0.00416	0.68569
Var_V35	0.00859	0.4026
Var_V36	-0.13311	0.0
Var_V37	-0.05771	0.0
Var_V38	0.10593	0.0
Var_V39	0.01397	0.17348
Var_V40	0.1113	0.0
Var_V41	-0.01565	0.12751
Var_V42	0.03	0.00347
Var_V43	-0.10038	0.0

Variable 22: Atypical houses [normalized by range method]	Pearson Correlation	p-value
Var_V23	-0.02173	0.03429
Var_V24	-0.01436	0.16199
Var_V25	0.0149	0.14682
Var_V26	0.01299	0.20579
Var_V27	-0.01264	0.2182
Var_V28	-0.01271	0.21588
Var_V29	-0.00998	0.33087
Var_V30	-0.03606	0.00044
Var_V31	0.03856	0.00017
Var_V32	-0.00897	0.38225
Var_V33	-0.01259	0.22007
Var_V34	0.00701	0.49494

Var_V35	-0.01265	0.21778
Var_V36	-0.0088	0.39118
Var_V37	0.03591	0.00047
Var_V38	-0.03127	0.00232
Var_V39	-0.00106	0.91761
Var_V40	0.04222	4e-05
Var_V41	-0.02046	0.04627
Var_V42	-0.00944	0.3579
Var_V43	-0.01197	0.24379

Variable 23: Average area per house [normalized by range method]	**Pearson Correlation**	**p-value**
Var_V24	-0.10739	0.0
Var_V25	0.16849	0.0
Var_V26	0.0621	0.0
Var_V27	0.00799	0.43646
Var_V28	-0.13225	0.0
Var_V29	0.20319	0.0
Var_V30	0.07686	0.0
Var_V31	-0.18793	0.0
Var_V32	0.09778	0.0
Var_V33	0.12356	0.0
Var_V34	0.16166	0.0
Var_V35	0.25824	0.0
Var_V36	-0.10109	0.0
Var_V37	0.09862	0.0
Var_V38	0.00716	0.48586
Var_V39	-0.0423	4e-05
Var_V40	-0.02166	0.0349
Var_V41	0.09127	0.0
Var_V42	0.0198	0.05383
Var_V43	0.13903	0.0

Variable 24: Percentage of old houses 1919 - 1971	**Pearson Correlation**	**p-value**
Var_V25	-0.96898	0.0
Var_V26	-0.16793	0.0
Var_V27	0.20769	0.0
Var_V28	0.20748	0.0
Var_V29	0.10551	0.0
Var_V30	-0.21487	0.0
Var_V31	0.13382	0.0
Var_V32	-0.04721	0.0
Var_V33	-0.15072	0.0
Var_V34	-0.01671	0.10366
Var_V35	-0.06716	0.0
Var_V36	0.22328	0.0
Var_V37	-0.05645	0.0
Var_V38	0.06505	0.0
Var_V39	-0.04573	1e-05
Var_V40	-0.06424	0.0
Var_V41	0.13256	0.0
Var_V42	-0.03588	0.00047
Var_V43	0.04754	0.0

Variable 25: Percentage of recent houses > 1972	**Pearson Correlation**	**p-value**

Var_V26	0.22188	0.0
Var_V27	-0.19031	0.0
Var_V28	-0.1896	0.0
Var_V29	-0.12107	0.0
Var_V30	0.21774	0.0
Var_V31	-0.12738	0.0
Var_V32	0.07246	0.0
Var_V33	0.1656	0.0
Var_V34	0.04002	0.0001
Var_V35	0.09058	0.0
Var_V36	-0.25582	0.0
Var_V37	0.06564	0.0
Var_V38	-0.04558	1e-05
Var_V39	0.04923	0.0
Var_V40	0.0821	0.0
Var_V41	-0.13951	0.0
Var_V42	0.05434	0.0
Var_V43	-0.0378	0.00023

Variable 26: Percentage of building with 1 or 2 roofs	Pearson Correlation	p-value
Var_V27	-0.95876	0.0
Var_V28	-0.76482	0.0
Var_V29	-0.35514	0.0
Var_V30	-0.04006	0.0001
Var_V31	0.24283	0.0
Var_V32	-0.17072	0.0
Var_V33	0.14808	0.0
Var_V34	-0.05892	0.0
Var_V35	-0.14107	0.0
Var_V36	-0.04099	7e-05
Var_V37	0.13906	0.0
Var_V38	-0.17946	0.0
Var_V39	0.25022	0.0
Var_V40	0.15771	0.0
Var_V41	-0.33422	0.0
Var_V42	-0.00174	0.86555
Var_V43	-0.25034	0.0

Variable 27: Percentage of condominium	Pearson Correlation	p-value
Var_V28	0.8024	0.0
Var_V29	0.3447	0.0
Var_V30	0.04432	2e-05
Var_V31	-0.24063	0.0
Var_V32	0.20398	0.0
Var_V33	-0.13391	0.0
Var_V34	0.08747	0.0
Var_V35	0.17154	0.0
Var_V36	0.0038	0.71131
Var_V37	-0.13134	0.0
Var_V38	0.20614	0.0
Var_V39	-0.25158	0.0
Var_V40	-0.14026	0.0
Var_V41	0.33338	0.0
Var_V42	0.02336	0.02288
Var_V43	0.26719	0.0

Variable 28: Percentage of building with > 3 apartment numbers	Pearson Correlation	p-value
Var_V29	0.31766	0.0
Var_V30	0.05681	0.0
Var_V31	-0.23626	0.0
Var_V32	0.21091	0.0
Var_V33	-0.18266	0.0
Var_V34	0.06997	0.0
Var_V35	0.13282	0.0
Var_V36	-0.01504	0.14284
Var_V37	-0.1518	0.0
Var_V38	0.22179	0.0
Var_V39	-0.28858	0.0
Var_V40	-0.11151	0.0
Var_V41	0.30216	0.0
Var_V42	0.06598	0.0
Var_V43	0.26758	0.0

Variable 29: Percentage of people with High education level	Pearson Correlation	p-value
Var_V30	-0.15424	0.0
Var_V31	-0.44122	0.0
Var_V32	0.10975	0.0
Var_V33	-0.05201	0.0
Var_V34	0.14617	0.0
Var_V35	0.11203	0.0
Var_V36	0.03446	0.00079
Var_V37	-0.00742	0.46967
Var_V38	0.02117	0.03924
Var_V39	-0.21178	0.0
Var_V40	-0.22731	0.0
Var_V41	0.37597	0.0
Var_V42	-0.10358	0.0
Var_V43	0.30373	0.0

Variable 30: Percentage of people with Medium education level	Pearson Correlation	p-value
Var_V31	-0.81861	0.0
Var_V32	0.14889	0.0
Var_V33	0.12136	0.0
Var_V34	0.16125	0.0
Var_V35	0.24026	0.0
Var_V36	-0.32955	0.0
Var_V37	-0.0019	0.8529
Var_V38	0.12126	0.0
Var_V39	-0.07934	0.0
Var_V40	0.01337	0.19292
Var_V41	0.05103	0.0
Var_V42	0.13367	0.0
Var_V43	0.07811	0.0

Variable 31: Percentage of people with Low or No education level	Pearson Correlation	p-value
Var_V32	-0.19902	0.0
Var_V33	-0.08	0.0
Var_V34	-0.23142	0.0
Var_V35	-0.28334	0.0
Var_V36	0.27929	0.0
Var_V37	0.00604	0.55605

Var_V38	-0.12244	0.0
Var_V39	0.19517	0.0
Var_V40	0.12	0.0
Var_V41	-0.26491	0.0
Var_V42	-0.06119	0.0
Var_V43	-0.24751	0.0

Variable 32: Percentage of commuters	Pearson Correlation	p-value
Var_V33	-0.36271	0.0
Var_V34	0.36873	0.0
Var_V35	0.14953	0.0
Var_V36	-0.42832	0.0
Var_V37	0.08387	0.0
Var_V38	0.12855	0.0
Var_V39	-0.00431	0.67447
Var_V40	-0.06082	0.0
Var_V41	0.12847	0.0
Var_V42	0.11969	0.0
Var_V43	0.1333	0.0

Variable 33: Percentage of inner-municipality commuters	Pearson Correlation	p-value
Var_V34	0.10445	0.0
Var_V35	0.12993	0.0
Var_V36	-0.18814	0.0
Var_V37	-0.09218	0.0
Var_V38	0.18351	0.0
Var_V39	-0.02455	0.01679
Var_V40	0.11539	0.0
Var_V41	0.08577	0.0
Var_V42	-0.01878	0.06735
Var_V43	-0.04599	1e-05

Variable 34: Percentage of employed	Pearson Correlation	p-value
Var_V35	0.13731	0.0
Var_V36	-0.35618	0.0
Var_V37	0.27372	0.0
Var_V38	0.25255	0.0
Var_V39	0.13186	0.0
Var_V40	0.0535	0.0
Var_V41	0.21766	0.0
Var_V42	0.14178	0.0
Var_V43	0.14036	0.0

Variable 35: Percentage of Students	Pearson Correlation	p-value
Var_V36	-0.18995	0.0
Var_V37	-0.06119	0.0
Var_V38	0.14743	0.0
Var_V39	-0.17356	0.0
Var_V40	-0.01216	0.23612
Var_V41	0.1968	0.0
Var_V42	0.04531	1e-05
Var_V43	0.12382	0.0

Variable 36: Percentage of Not employed (retired or other condition)	Pearson Correlation	p-value
Var_V37	-0.14462	0.0
Var_V38	-0.1925	0.0
Var_V39	-0.11657	0.0
Var_V40	-0.08242	0.0
Var_V41	-0.07446	0.0
Var_V42	-0.10909	0.0
Var_V43	-0.04025	9e-05

Variable 37: Percentage of businessman of freelancer	Pearson Correlation	p-value
Var_V38	-0.71752	0.0
Var_V39	0.49305	0.0
Var_V40	-0.06598	0.0
Var_V41	-0.29586	0.0
Var_V42	0.03066	0.00282
Var_V43	0.01688	0.10016

Variable 38: Percentage of salaried worker	Pearson Correlation	p-value
Var_V39	-0.37967	0.0
Var_V40	0.20778	0.0
Var_V41	0.45307	0.0
Var_V42	0.13272	0.0
Var_V43	0.08458	0.0

Variable 39: Percentage of employers in Agriculture	Pearson Correlation	p-value
Var_V40	-0.20238	0.0
Var_V41	-0.37876	0.0
Var_V42	-0.31676	0.0
Var_V43	-0.22427	0.0

Variable 40: Percentage of employers in Industry	Pearson Correlation	p-value
Var_V41	-0.26861	0.0
Var_V42	-0.16033	0.0
Var_V43	-0.17064	0.0

Variable 41: Percentage of employers in Public Services	Pearson Correlation	p-value
Var_V42	-0.20305	0.0
Var_V43	0.06654	0.0

Variable 42: Percentage of employers in Trade, Restaurant, Transport, Communication	Pearson Correlation	p-value
Var_V43	-0.05986	0.0

Variable 43: Percentage of employers in Financial intermediation and business	Pearson Correlation	p-value
Var_V43	1.0	0.0

Notwithstanding the existence of a few highly correlated variables, both positively or negatively, such as the 'Percentage of people with Medium education level' (Variable 30) with the 'Percentage of people with Low or No education level' (Variable 31) (-0.81861 correlation value and p-value 0.0), or the 'Percentage of condominium' (Variable 27) with the 'Percentage of building with > 3 apartment numbers' (Variable 28) (0.8024 correlation value and p-value 0.0), the variables set is not modified. The choice to keep all the 43 variables is due to the opportunity arising from the highly correlated variables to better discriminate among groups when using the k-means clustering method for the geodemographic classification. As a matter of fact, these correlated variables present a notable descriptive and predictive power that was used as an advantage for partitioning.

ANNEX 2

GEODEMOGRAPHIC CLASSIFICATION DESCRIPTIVE SHEETS

A2.1 Introduction

The descriptive sheets are created in order to provide the results of quantitative and qualitative analyses conducted on the geodemographic classification of Sardinia census tracts. The analysis of the first hierarchical level groups is conducted by comparing each group variables means with the means of the variables at the regional scale. Similarly, the second hierarchical level sub-groups are evaluated considering the sub-groups means of variables with their father group's means. The evaluation is conducted by means of radial plots and histograms, which stress the main distinguishing features of each group and sub-group. In addition, the variables fostering the identification of a descriptive label are highlighted in terms of positive or negative deviation from their mean.

A2.2 First hierarchical level groups

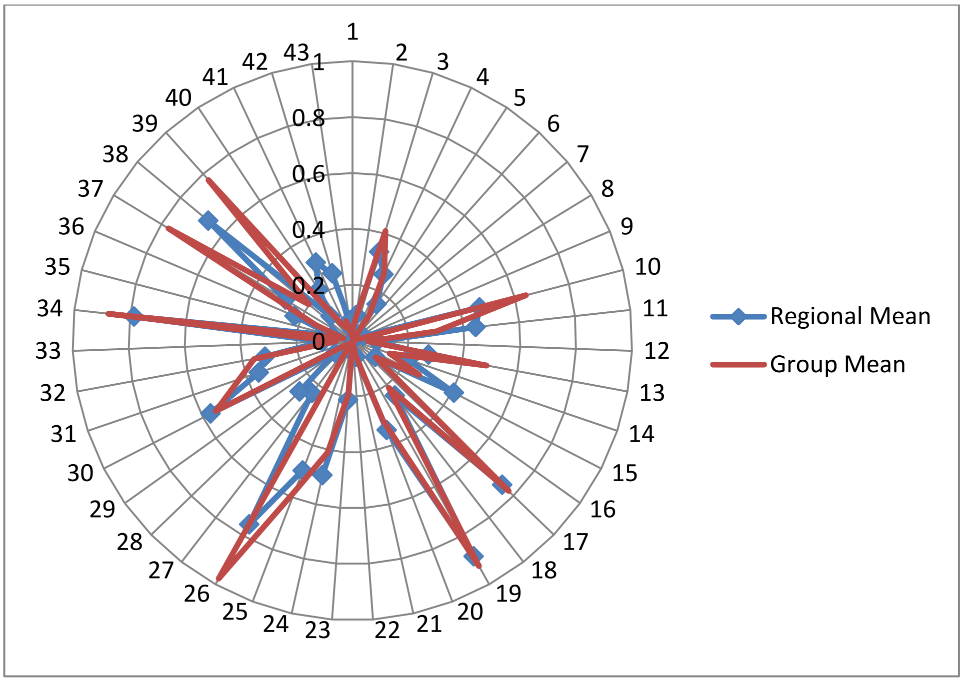
The first hierarchical level of the Sardinia geodemographic classification consists of 6 groups, which are singularly described by depicting in the radial plot both the variables means at the regional scale and the inherent group means. At the same time, the provided histogram identifies the major differences allowing the identification of the group's characteristic features. Finally, a short textual description is provided on the base of obtained findings. The used variables and the descriptive sheets are following provided.

VARIABLE	DESCRIPTION	CATEGORY
V1	Percentage of resident population aged 0-4 years	Demographics
V2	Percentage of resident population aged 5-14 years	Demographics
V3	Percentage of resident population aged 25-44 years	Demographics
V4	Percentage of resident population aged 45-64 years	Demographics
V5	Percentage of resident population aged > 65 years	Demographics
V6	Percentage of resident population Afrikans	Demographics
V7	Percentage of resident population Asian	Demographics
V8	Percentage of resident population Caucasic of Hispanic	Demographics
V9	Density Pop/sqKm [normalized by range method]	Demographics
V10	Percentage of unmarried	Household composition
V11	Percentage of married	Household composition
V12	Percentage of separated, divorced or widowed	Household composition
V13	Percentage of single person household	Household composition
V14	Percentage of couple household	Household composition
V15	Percentage of household of 3-4 persons	Household composition
V16	Percentage of household > 5 persons	Household composition
V17	Percentage of house owners	Housing Typology
V18	Percentage of house tenants	Housing Typology
V19	Percentage of house with heating	Housing Typology
V20	Average rooms per house [normalized by range method]	Housing Typology
V21	Average persons per room [normalized by range method]	Housing Typology
V22	Atypical houses [normalized by range method]	Housing Typology
V23	Average area per house [normalized by range method]	Housing Typology
V24	Percentage of old houses 1919 - 1971	Housing Typology
V25	Percentage of recent houses > 1972	Housing Typology
V26	Percentage of building with 1 or 2 roofs	Housing Typology
V27	Percentage of condominium	Housing Typology
V28	Percentage of building with > 3 apartment numbers	Housing Typology
V29	Percentage of people with High education level	Socio economic
V30	Percentage of people with Medium education level	Socio economic
V31	Percentage of people with Low or No education level	Socio economic
V32	Percentage of commuters	Socio economic
V33	Percentage of inner-municipality commuters	Socio economic
V34	Percentage of employed	Socio economic
V35	Percentage of Students	Employment condition
V36	Percentage of Not employed (retired or other condition)	Employment condition
V37	Percentage of businessman of freelancer	Employment condition
V38	Percentage of salaried worker	Employment condition
V39	Percentage of employers in Agriculture	Employment condition
V40	Percentage of employers in Industry	Employment condition
V41	Percentage of employers in Public Services	Employment condition
V42	Percentage of employers in Trade, Restaurant, Transport, Communication	Employment condition
V43	Percentage of employers in Financial intermediation and business	Employment condition

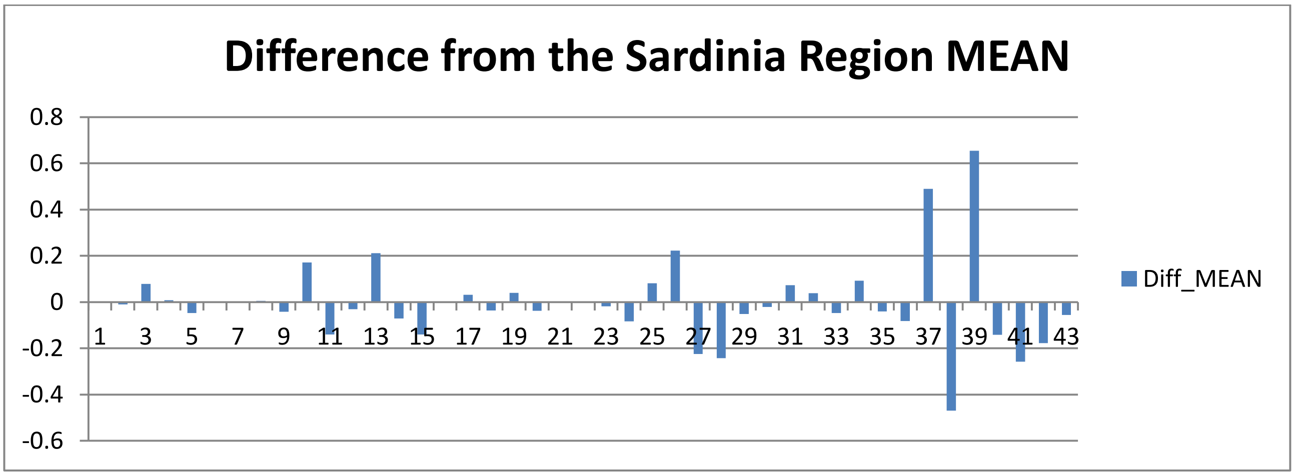
Table 41. SMGI Analytics Geodemographics: selected socio-economic variables.

A2.2.1 Group 1 sheet

The sheet describes the Group 1 main characteristics as obtained from the 43 variables measurements.



Radial plot Group 1. Variables measurements.



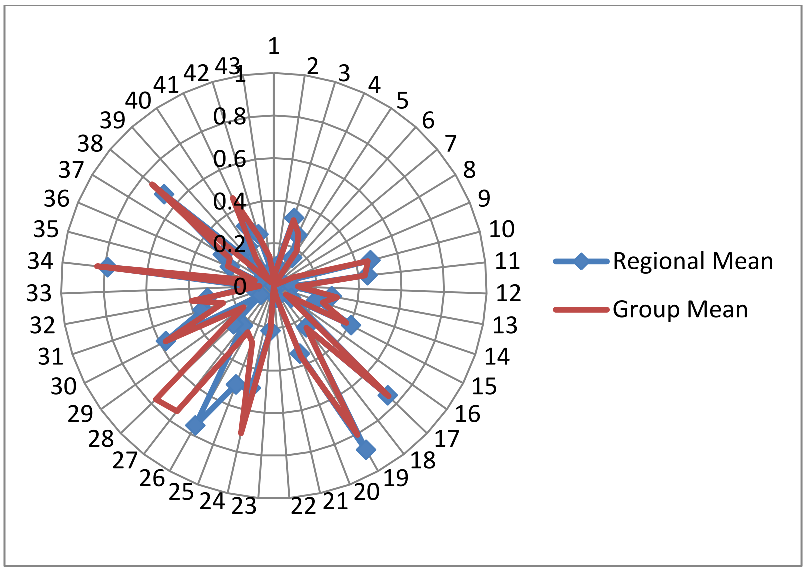
V13	Single person household	Household composition	POS
V26	Percentage of building with 1 or 2 roofs	Housing Typology	POS
V27	Percentage of condominium	Housing Typology	NEG
V28	Percentage of building with > 3 apartment numbers	Housing Typology	NEG
V37	Percentage of businessman of freelancer	Employment condition	POS
V38	Percentage of salaried worker	Employment condition	NEG
V39	Percentage of employers in Agricolture	Employment condition	POS
V41	Percentage of employers in Public Services	Employment condition	NEG

Histogram Group 1. Variables differences from regional means.

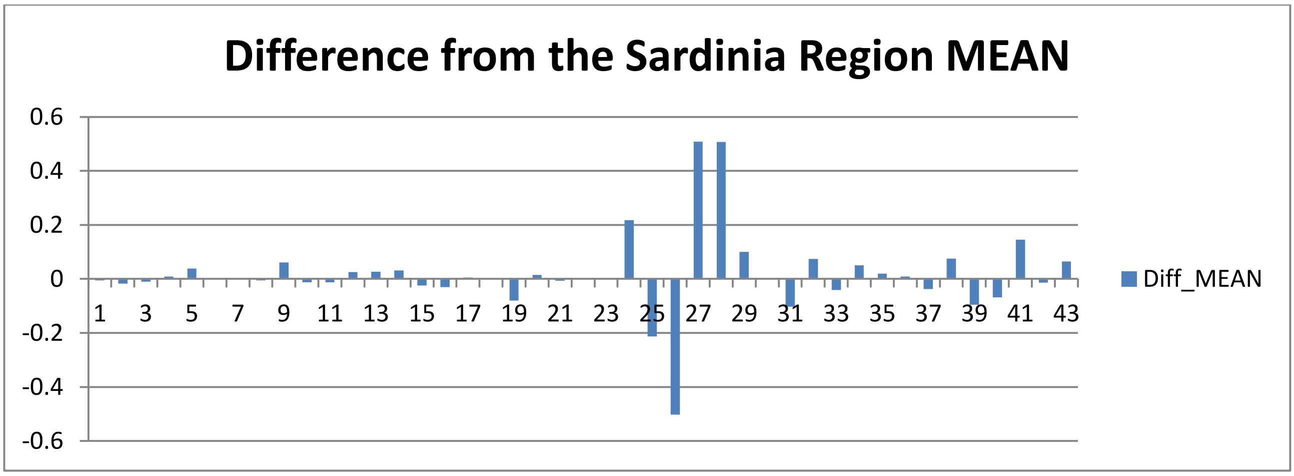
LABEL: Single farmers in independent buildings.

A2.2.2 Group 2 sheet

The sheet describes the Group 2 main characteristics as obtained from the 43 variables measurements.



Radial plot Group 2. Variables measurements.



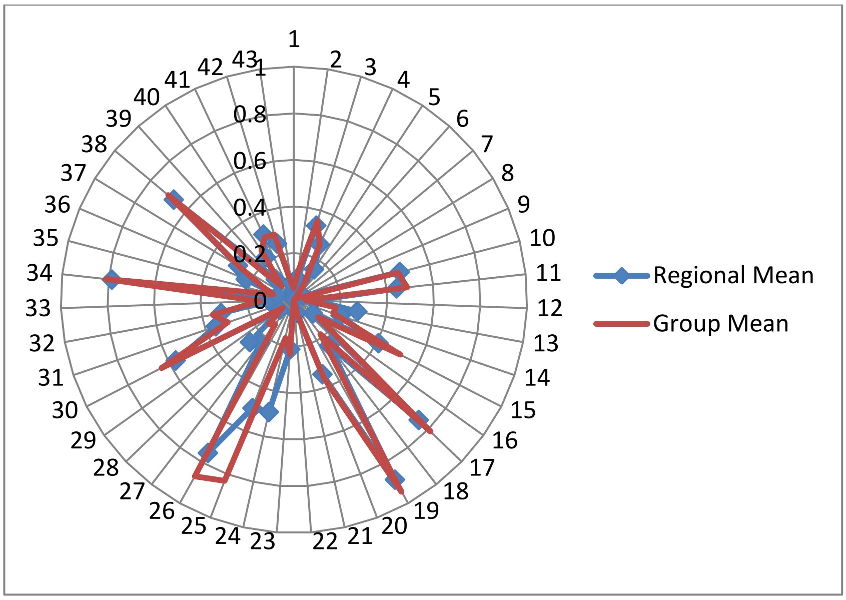
V24	Percentage of old houses 1919 - 1971	**Housing Typology**	**POS**
V25	Percentage of recent houses > 1972	**Housing Typology**	**NEG**
V26	Percentage of building with 1 or 2 roofs	**Housing Typology**	**NEG**
V27	Percentage of condominium	**Housing Typology**	**POS**
V28	Percentage of building with > 3 apartment numbers	**Housing Typology**	**POS**

Histogram Group 2. Variables differences from regional means.

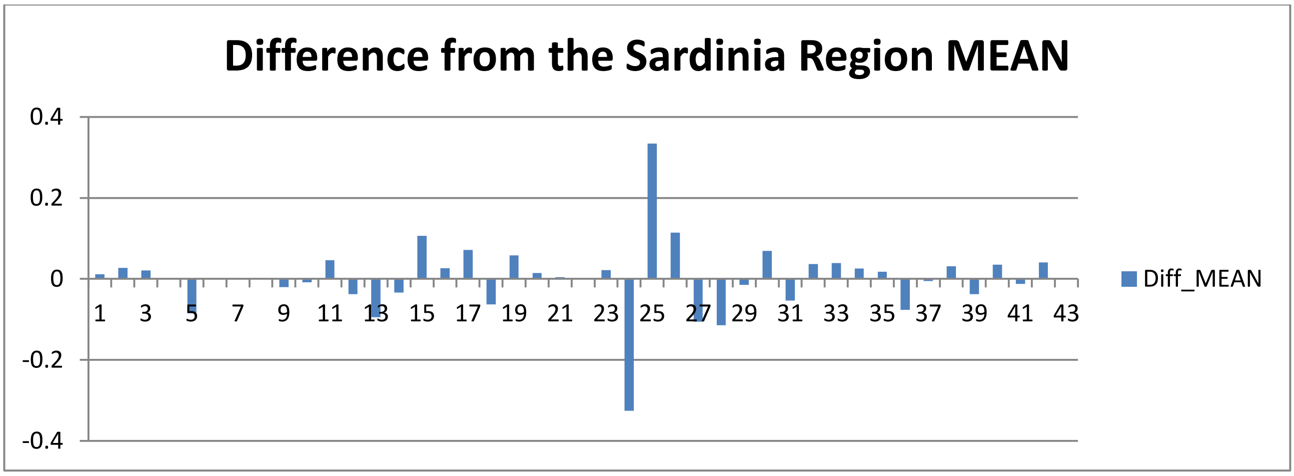
LABEL: Residents of condominium and multi-apartment buildings.

A2.2.3 Group 3 sheet

The sheet describes the Group 3 main characteristics as obtained from the 43 variables measurements.



Radial plot Group 3. Variables measurements.



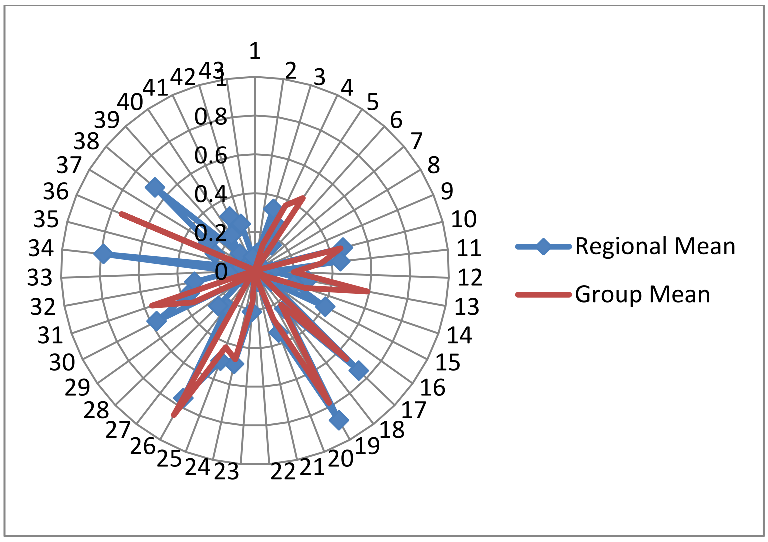
V24	Percentage of old houses 1919 - 1971	**Housing Typology**	**NEG**
V25	Percentage of recent houses > 1972	**Housing Typology**	**POS**

Histogram Group 3. Variables differences from regional means.

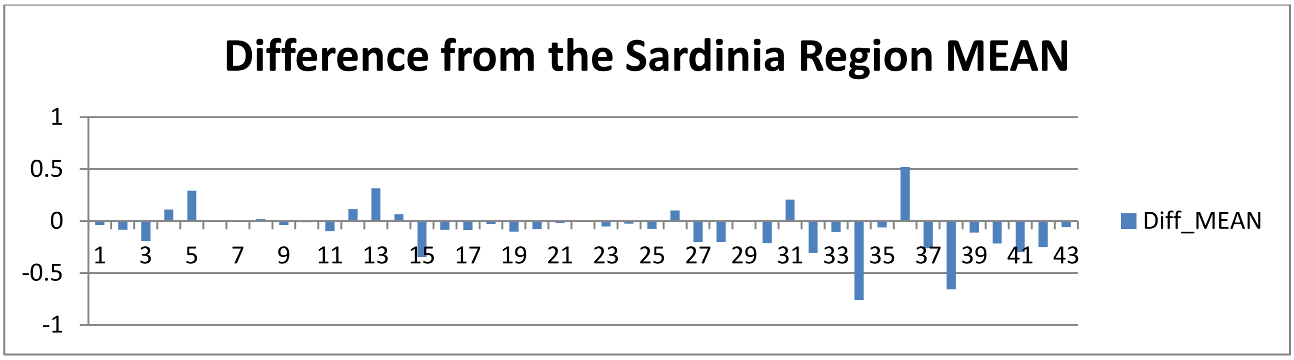
LABEL: Recent buildings residents.

A2.2.4 Group 4 sheet

The sheet describes the Group 4 main characteristics as obtained from the 43 variables measurements.



Radial plot Group 4. Variables measurements.



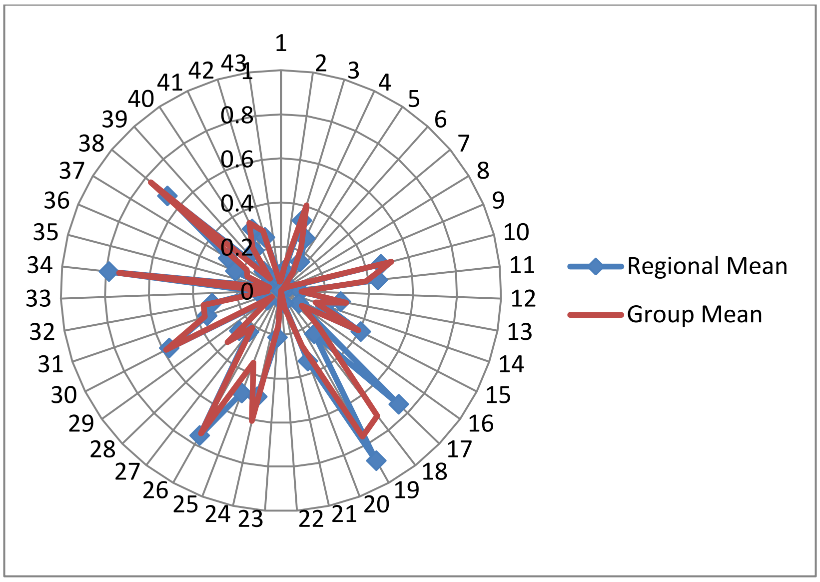
V5	Percentage of resident population aged > 65 years	**Demographics**	**POS**
V13	Single person household	**Household composition**	**POS**
V15	Household of 3-4 persons	**Household composition**	**NEG**
V27	Percentage of condominium	**Housing Typology**	**NEG**
V28	Percentage of building with > 3 apartment numbers	**Housing Typology**	**NEG**
V30	Percentage of people with Medium education level	**Socio economic**	**NEG**
V31	Percentage of people with Low or No education level	**Socio economic**	**POS**
V32	Percentage of commuters	**Socio economic**	**NEG**
V34	Percentage of employed	**Socio economic**	**NEG**
V36	Percentage of Not employed (retired or other condition)	**Employment condition**	**POS**
V37	Percentage of businessman of freelancer	**Employment condition**	**NEG**
V38	Percentage of salaried worker	**Employment condition**	**NEG**
V40	Percentage of employers in Industry	**Employment condition**	**NEG**
V41	Percentage of employers in Public Services	**Employment condition**	**NEG**
V42	Percentage of employers in Trade, Restaurant, Transport, Communication	**Employment condition**	**NEG**

Histogram Group 4. Variables differences from regional means.

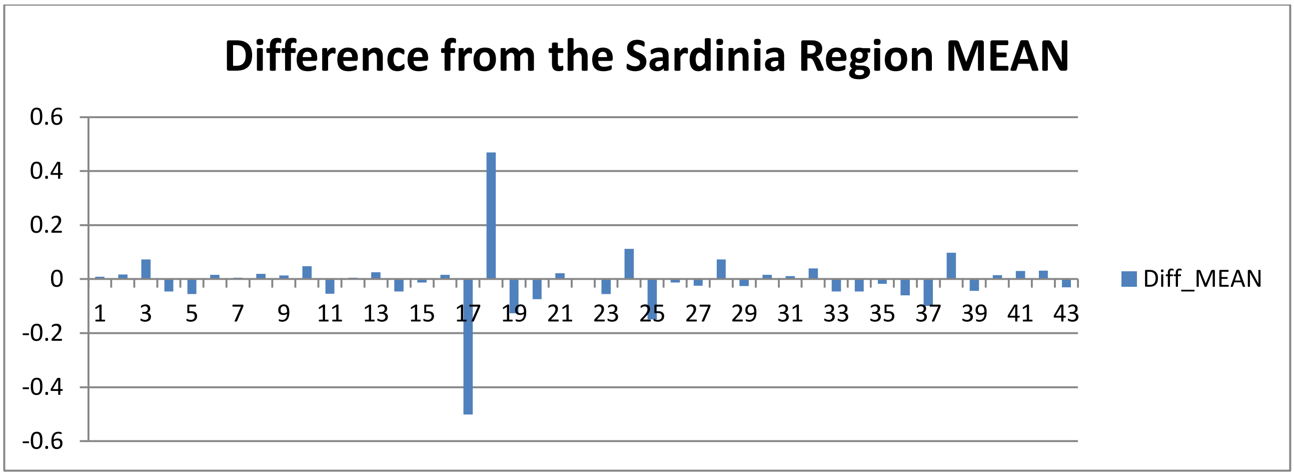
LABEL: Retired in small buildings.

A2.2.5 Group 5 sheet

The sheet describes the Group 5 main characteristics as obtained from the 43 variables measurements.



Radial plot Group 5. Variables measurements.



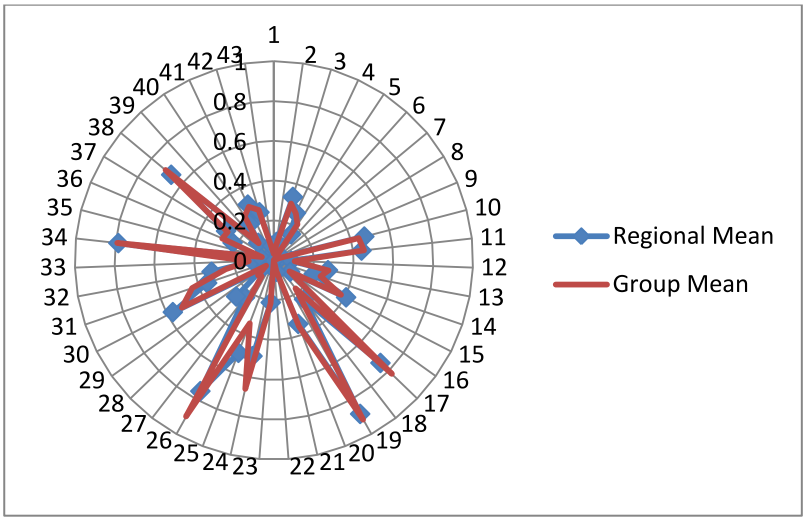
V17	House owners	**Housing Typology**	**NEG**
V18	House tenants	**Housing Typology**	**POS**

Histogram Group 5. Variables differences from regional means.

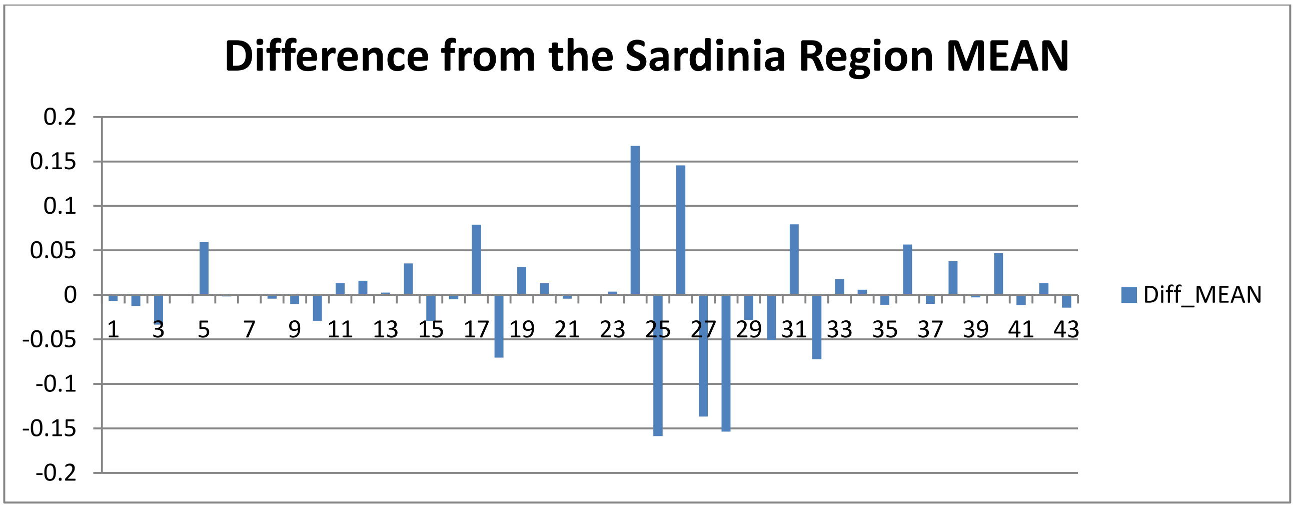
LABEL: Tenants.

A2.2.6 Group 6 sheet

The sheet describes the Group 6 main characteristics as obtained from the 43 variables measurements.



Radial plot Group 6. Variables measurements.



V24	Percentage of old houses 1919 - 1971	**Housing Typology**	**POS**
V25	Percentage of recent houses > 1972	**Housing Typology**	**NEG**
V26	Percentage of building with 1 or 2 roofs	**Housing Typology**	**POS**
V27	Percentage of condominium	**Housing Typology**	**NEG**
V28	Percentage of building with > 3 apartment numbers	**Housing Typology**	**NEG**

Histogram Group 6. Variables differences from regional means.

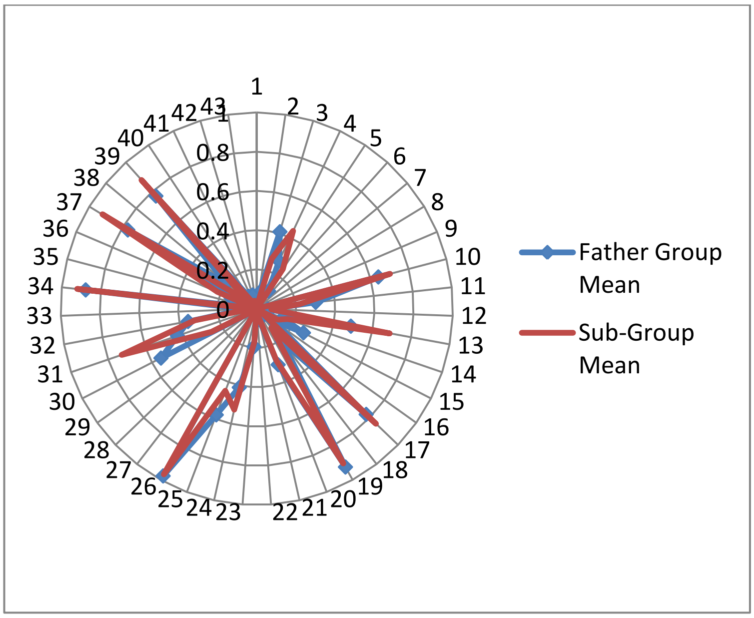
LABEL: Old buildings residents.

A2.3 Second hierarchical level groups

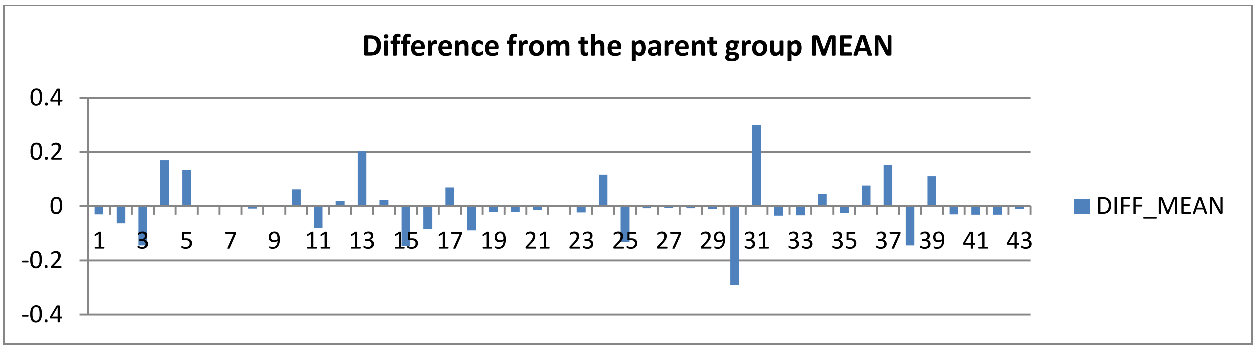
The second hierarchical level of the Sardinia geodemographic classification consists of 18 groups, which are singularly described by depicting in the radial plot both the variables means of the origin first level group and the inherent sub-group means. At the same time, the provided histogram identifies the major differences allowing the identification of the sub-group's characteristic features. Finally, a short textual description is provided on the base of obtained findings. The descriptive sheets are following provided.

A2.3.1 Sub-group 1.1 sheet [sub-group 1]

The sheet describes the sub-group 1.1 (s-gr.1) main features as obtained from the variables measures.



Radial plot Sub-Group 1.1.



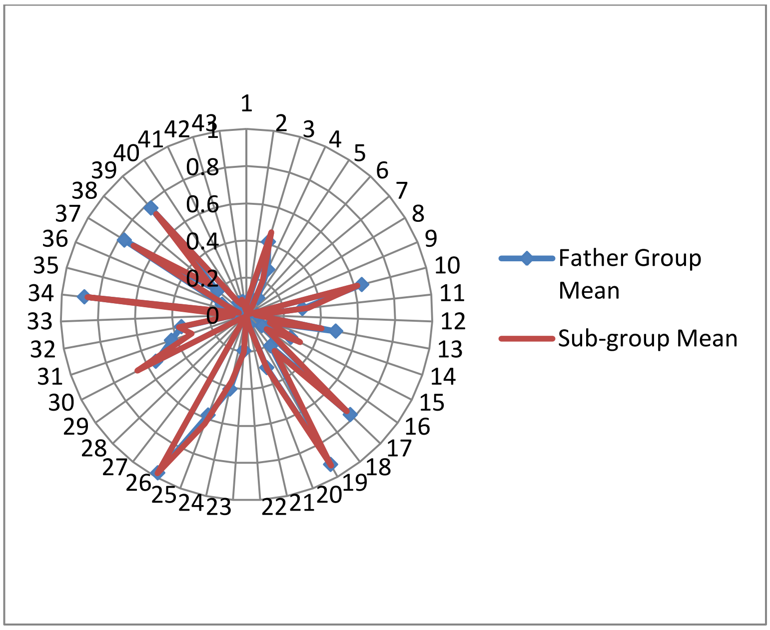
V13	Single person household	Household composition	**POS**
V30	Percentage of people with Medium education level	Socio economic	**NEG**
V31	Percentage of people with Low or No education level	Socio economic	**POS**
V4	Percentage of resident population aged 45-64 years	Demographics	**POS**
V37	Percentage of businessman of freelancer	Employment condition	**POS**

Histogram Sub-group 1.1. Variables differences from 1st level Group 1.

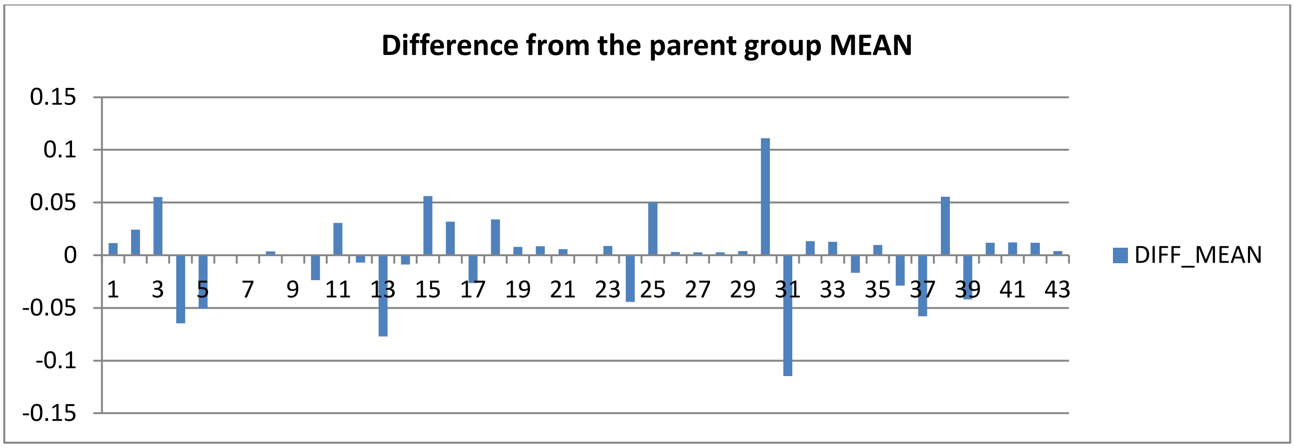
LABEL: Elder single farmers.

A2.3.2 Sub-group 1.2 sheet [sub-group 2]

The sheet describes the sub-group 1.2 (s-gr.2) main features as obtained from the variables measures.



Radial plot Sub-Group 1.2.



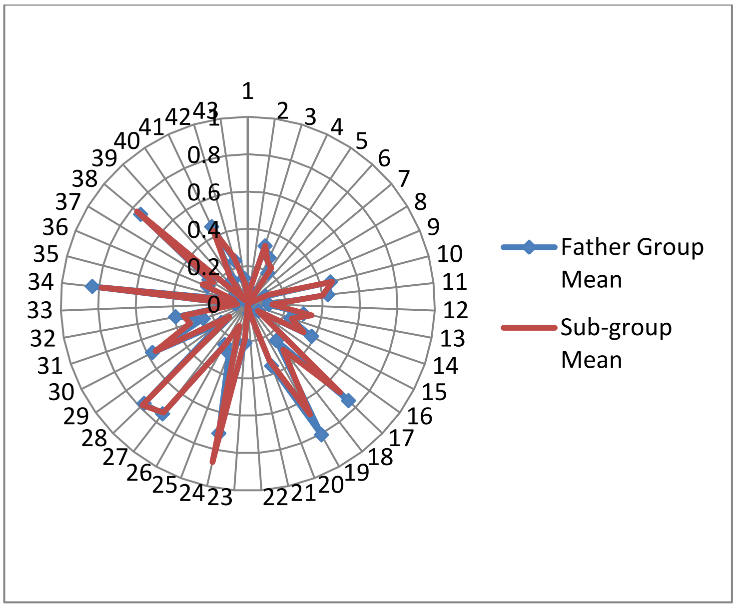
V30	Percentage of people with Medium education level	Socio economic	**POS**
V31	Percentage of people with Low or No education level	Socio economic	**NEG**
V13	Single person household	Household composition	**NEG**
V3	Percentage of resident population aged 25-44 years	Demographics	**POS**
V4	Percentage of resident population aged 45-64 years	Demographics	**NEG**
V5	Percentage of resident population aged > 65 years	Demographics	**NEG**
V15	Household of 3-4 persons	Household composition	**POS**
V25	Percentage of recent houses > 1972	Housing Typology	**POS**
V37	Percentage of businessman of freelancer	Employment condition	**NEG**
V38	Percentage of salaried worker	Employment condition	**POS**

Histogram Sub-group 1.2. Variables differences from 1st level Group 1.

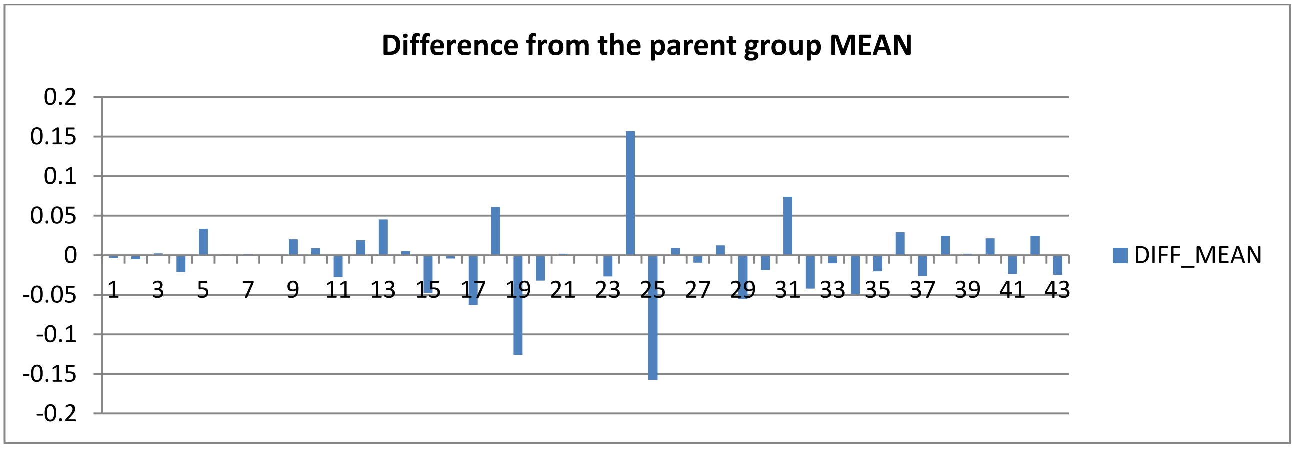
LABEL: Young single farmers.

A2.3.3 Sub-group 2.1 sheet [sub-group 3]

The sheet describes the sub-group 2.1 (s-gr.3) main features as obtained from the variables measures.



Radial plot Sub-Group 2.1.



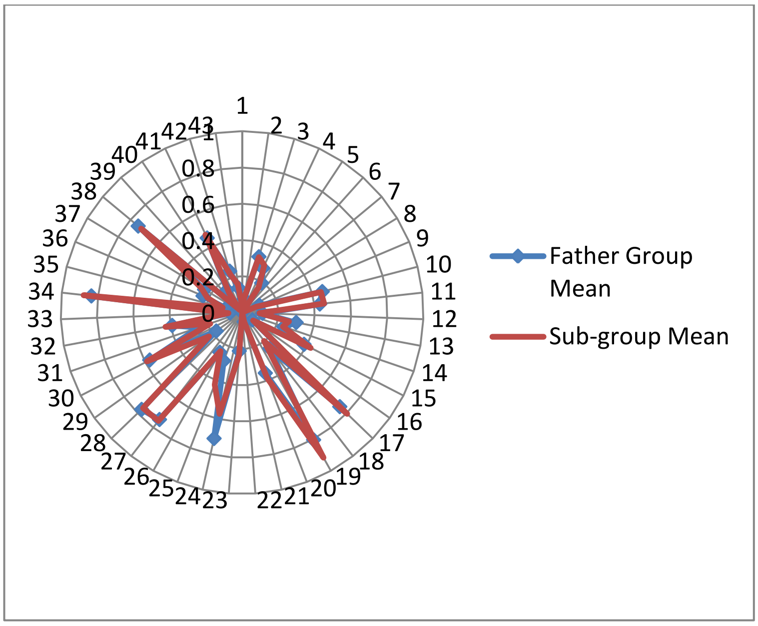
V24	Percentage of old houses 1919 - 1971	Housing Typology	**POS**
V25	Percentage of recent houses > 1972	Housing Typology	**NEG**
V19	Percentage of house with heating	Housing Typology	**NEG**
V17	House owners	Housing Typology	**NEG**
V18	House tenants	Housing Typology	**POS**
V29	Percentage of people with High education level	Socio economic	**NEG**
V31	Percentage of people with Low or No education level	Socio economic	**POS**

Histogram Sub-group 2.1. Variables differences from 1st level Group 2.

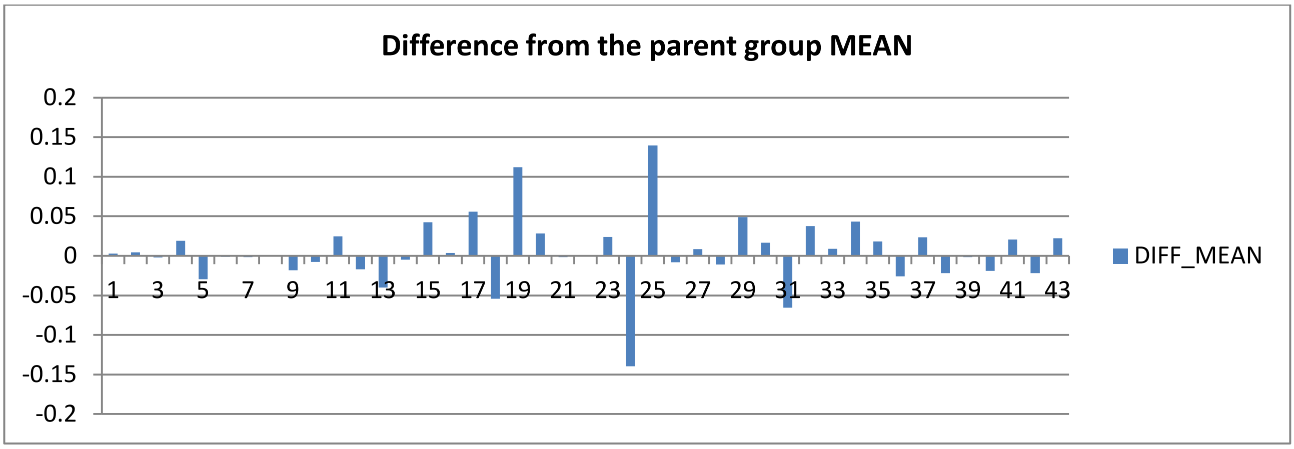
LABEL: Elder blue-collars.

A2.3.4 Sub-group 2.2 sheet [sub-group 4]

The sheet describes the sub-group 2.2 (s-gr.4) main features as obtained from the variables measures.



Radial plot Sub-Group 2.2.



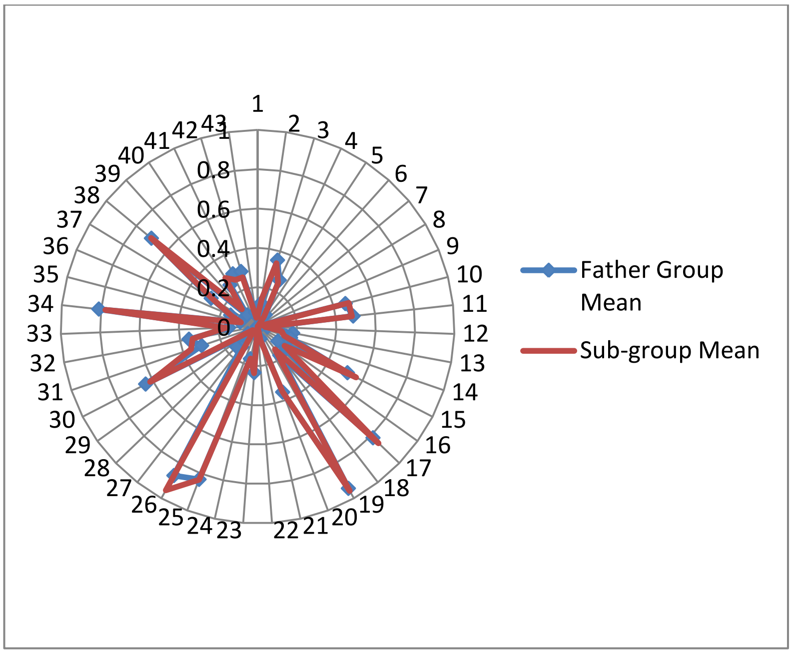
V19	Percentage of house with heating	Housing Typology	**POS**
V24	Percentage of old houses 1919 - 1971	Housing Typology	**NEG**
V25	Percentage of recent houses > 1972	Housing Typology	**POS**
V17	House owners	Housing Typology	**POS**
V18	House tenants	Housing Typology	**NEG**
V31	Percentage of people with Low or No education level	Socio economic	**NEG**

Histogram Sub-group 2.2. Variables differences from 1st level Group 2.

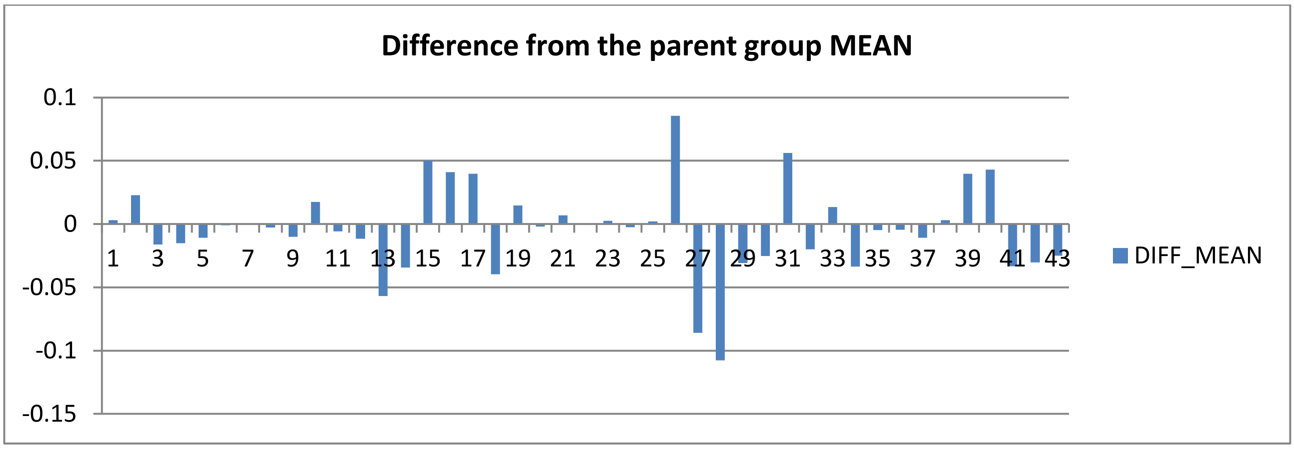
LABEL: Middle-class families.

A2.3.5 Sub-group 3.1 sheet [sub-group 5]

The sheet describes the sub-group 3.1 (s-gr.5) main features as obtained from the variables measures.



Radial plot Sub-Group 3.1.



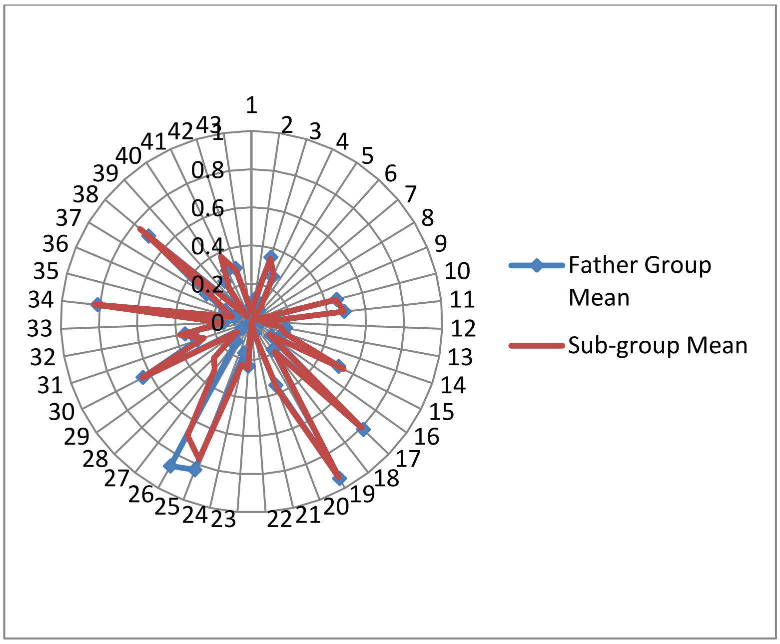
V28	Percentage of building with > 3 apartment numbers	Housing Typology	**NEG**
V26	Percentage of building with 1 or 2 roofs	Housing Typology	**POS**
V27	Percentage of condominium	Housing Typology	**NEG**
V13	Single person household	Household composition	**NEG**
V15	Household of 3-4 persons	Household composition	**POS**
V31	Percentage of people with Low or No education level	Socio economic	**POS**

Histogram Sub-group 3.1. Variables differences from 1st level Group 3.

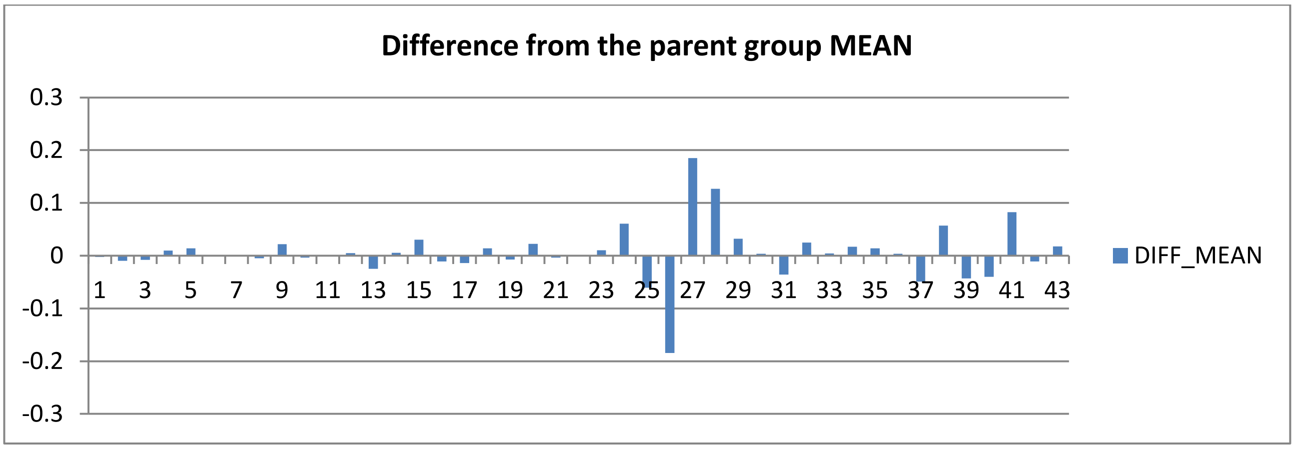
LABEL: Young Blue-collars families.

A2.3.6 Sub-group 3.2 sheet [sub-group 6]

The sheet describes the sub-group 3.2 (s-gr.6) main features as obtained from the variables measures.



Radial plot Sub-Group 3.2.



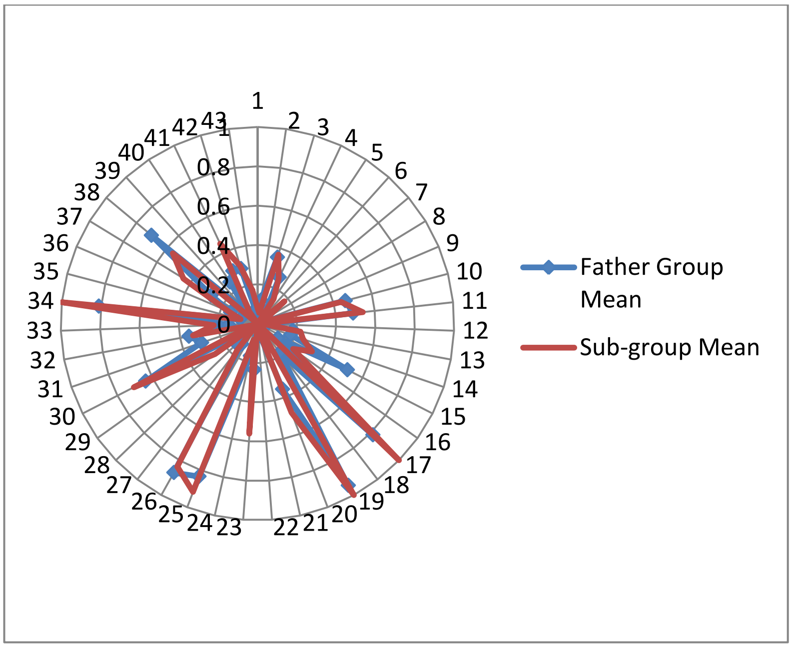
V26	Percentage of building with 1 or 2 roofs	Housing Typology	**NEG**
V27	Percentage of condominium	Housing Typology	**POS**
V28	Percentage of building with > 3 apartment numbers	Housing Typology	**POS**
V41	Percentage of employers in Public Services	Employment condition	**POS**

Histogram Sub-group 3.2. Variables differences from 1st level Group 3.

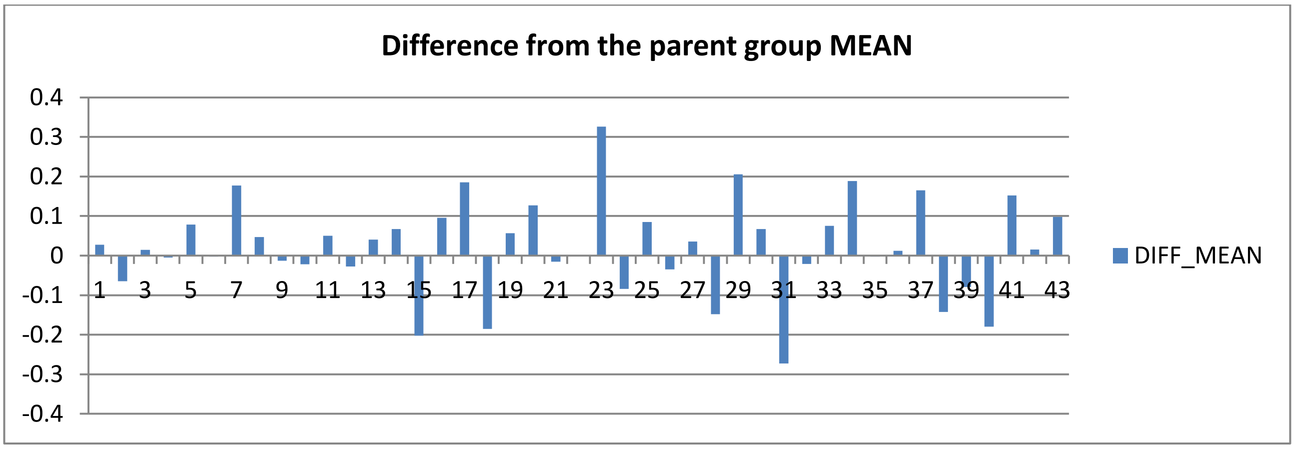
LABEL: Young prosperous families.

A2.3.7 Sub-group 3.3 sheet [sub-group 7]

The sheet describes the sub-group 3.3 (s-gr.7) main features as obtained from the variables measures.



Radial plot Sub-Group 3.3.



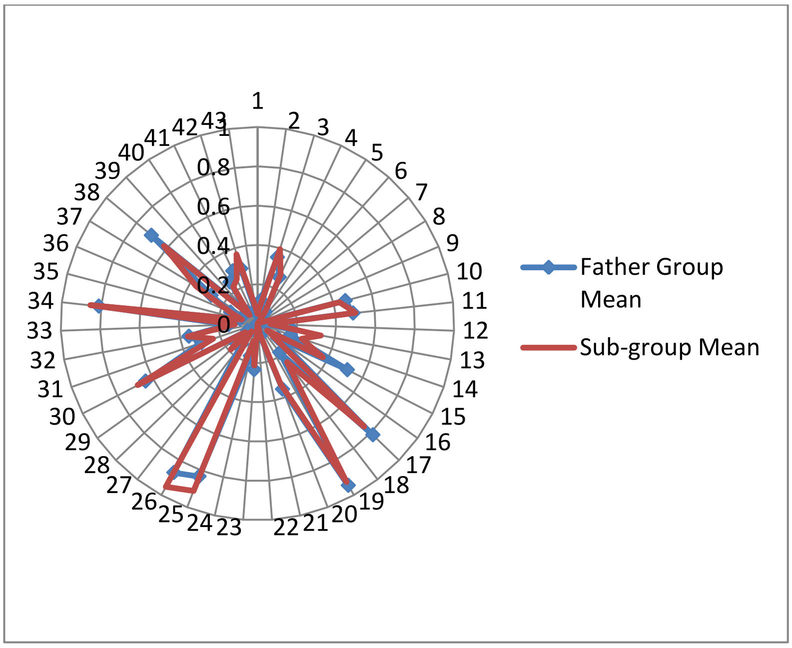
V15	Household of 3-4 persons	Household composition	**NEG**
V23	Average area per house [normalized by range method]	Housing Typology	**POS**
V29	Percentage of people with High education level	Socio economic	**POS**
V31	Percentage of people with Low or No education level	Socio economic	**NEG**
V7	Percentage of resident population Asian	Demographics	**POS**
V17	House owners	Housing Typology	**POS**
V18	House tenants	Housing Typology	**NEG**
V34	Percentage of employed	Socio economic	**POS**
V37	Percentage of businessman of freelancer	Employment condition	**POS**
V40	Percentage of employers in Industry	Employment condition	**NEG**
V41	Percentage of employers in Public Services	Employment condition	**POS**

Histogram Sub-group 3.3. Variables differences from 1st level Group 3.

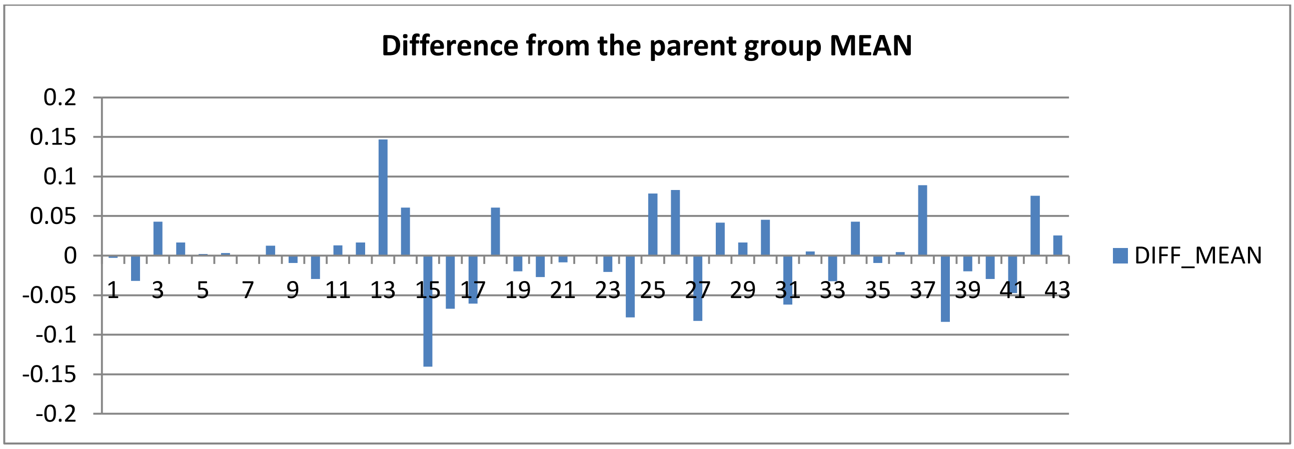
LABEL: Young entrepreneurs families.

A2.3.8 Sub-group 3.4 sheet [sub-group 8]

The sheet describes the sub-group 3.4 (s-gr.8) main features as obtained from the variables measures.



Radial plot Sub-Group 3.4.



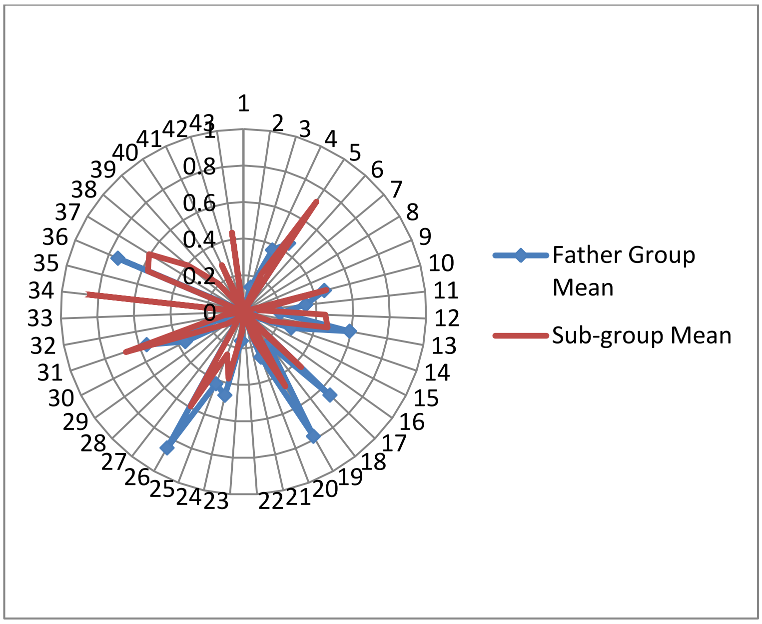
V13	Single person household	Household composition	**POS**
V15	Household of 3-4 persons	Household composition	**NEG**
V24	Percentage of old houses 1919 - 1971	Housing Typology	**NEG**
V25	Percentage of recent houses > 1972	Housing Typology	**POS**
V26	Percentage of building with 1 or 2 roofs	Housing Typology	**POS**
V27	Percentage of condominium	Housing Typology	**NEG**
V37	Percentage of businessman of freelancer	Employment condition	**POS**
V38	Percentage of salaried worker	Employment condition	**NEG**
V42	Percentage of employers in Trade, Restaurant, Transport, Communication	Employment condition	**POS**

Histogram Sub-group 3.4. Variables differences from 1st level Group 3.

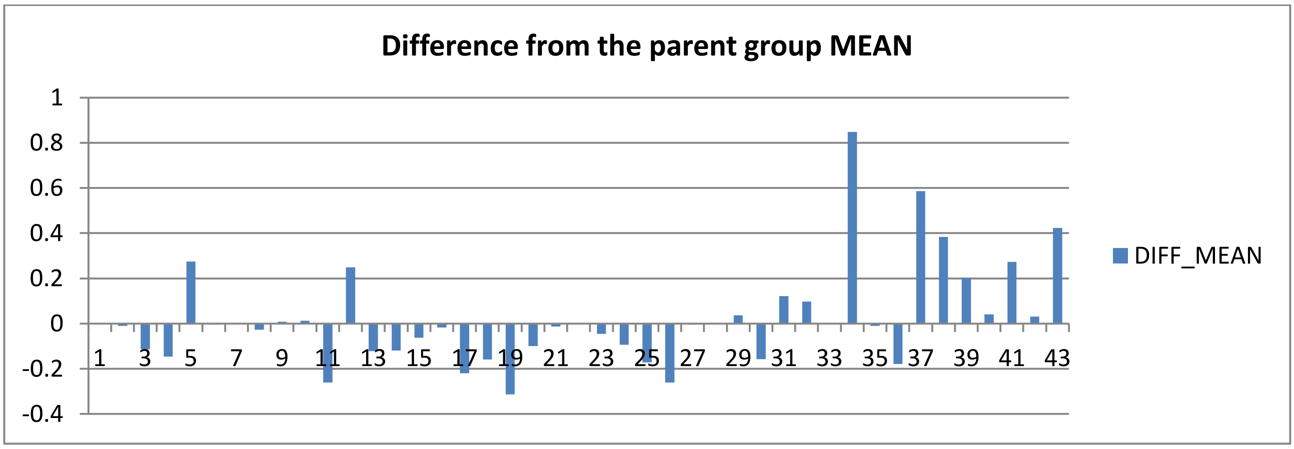
LABEL: Tradesman families.

A2.3.9 Sub-group 4.1 sheet [sub-group 9]

The sheet describes the sub-group 4.1 (s-gr.9) main features as obtained from the variables measures.



Radial plot Sub-Group 4.1.



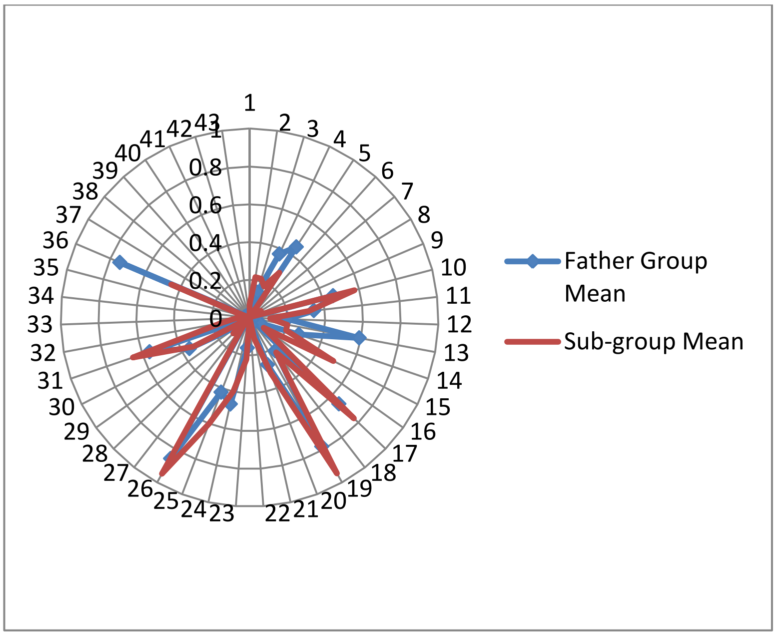
V5	Percentage of resident population aged > 65 years	Demographics	**POS**
V11	Married	Household composition	**NEG**
V12	Separated/Divorced/Widowed	Household composition	**POS**
V17	House owners	Housing Typology	**NEG**
V19	Percentage of house with heating	Housing Typology	**NEG**
V26	Percentage of building with 1 or 2 roofs	Housing Typology	**NEG**
V34	Percentage of employed	Socio economic	**POS**
V37	Percentage of businessman of freelancer	Employment condition	**POS**
V38	Percentage of salaried worker	Employment condition	**POS**
V39	Percentage of employers in Agricolture	Employment condition	**POS**
V41	Percentage of employers in Public Services	Employment condition	**POS**
V43	Percentage of employers in Financial intermediation and business	Employment condition	**POS**

Histogram Sub-group 4.1. Variables differences from 1st level Group 4.

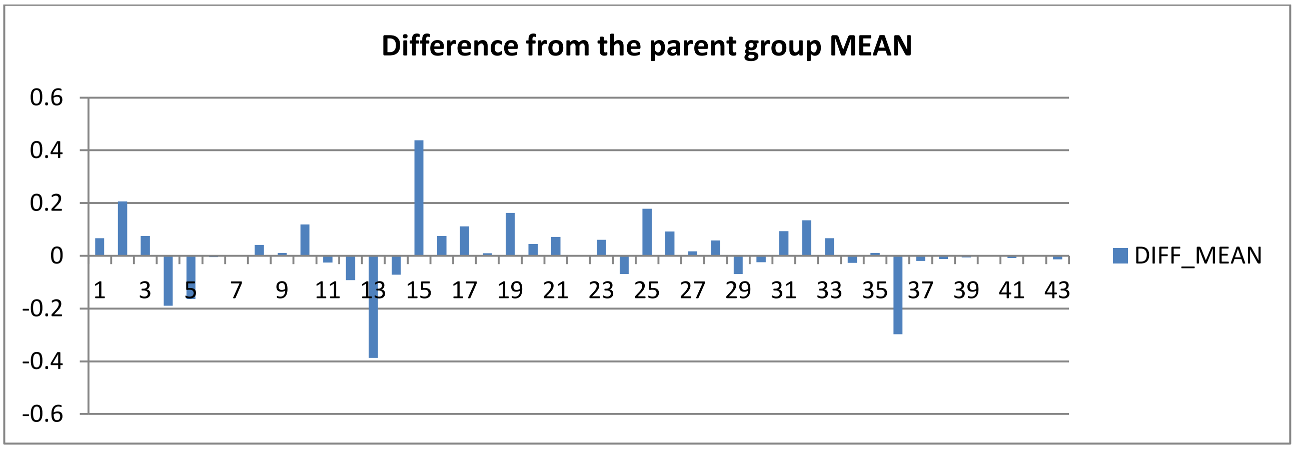
LABEL: Retired businessmen.

A2.3.10Sub-group 4.2 sheet [sub-group 10]

The sheet describes the sub-group 4.2 (s-gr.10) main features as obtained from the variables measures.



Radial plot Sub-Group 4.2.



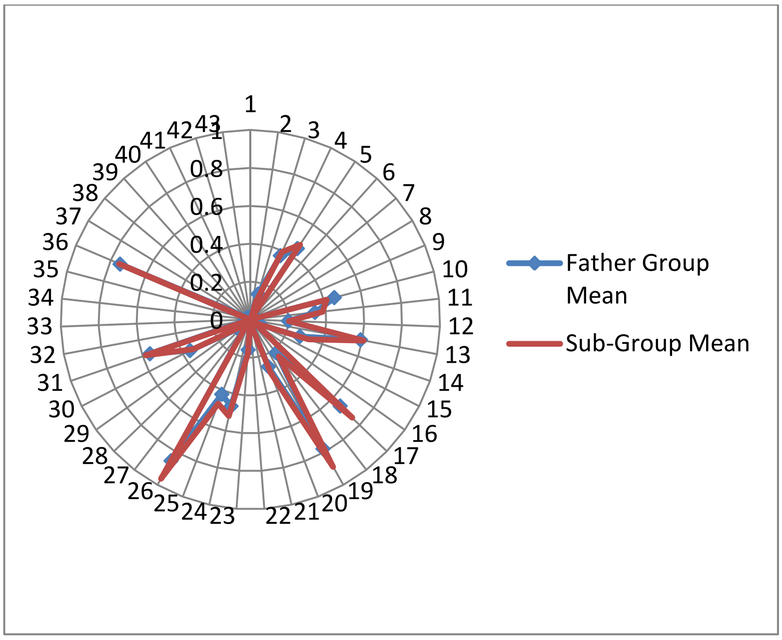
V2	Percentage of resident population aged 5-14 years	Demographics	**POS**
V13	Single person household	Household composition	**NEG**
V15	Household of 3-4 persons	Household composition	**POS**
V36	Percentage of Not employed (retired or other condition)	Employment condition	**NEG**
V4	Percentage of resident population aged 45-64 years	Demographics	**NEG**
V5	Percentage of resident population aged > 65 years	Demographics	**NEG**
V19	Percentage of house with heating	Housing Typology	**POS**
V25	Percentage of recent houses > 1972	Housing Typology	**POS**

Histogram Sub-group 4.2. Variables differences from 1st level Group 4.

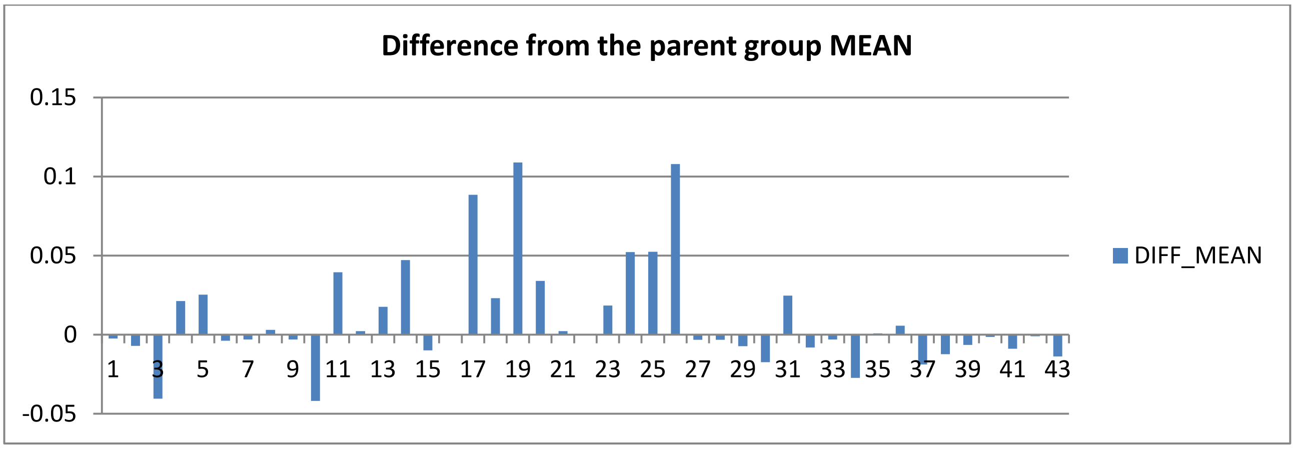
LABEL: Low-education retired on-a-budget.

A2.3.11 Sub-group 4.3 sheet [sub-group 11]

The sheet describes the sub-group 4.3 (s-gr.11) main features as obtained from the variables measures.



Radial plot Group 4.3.



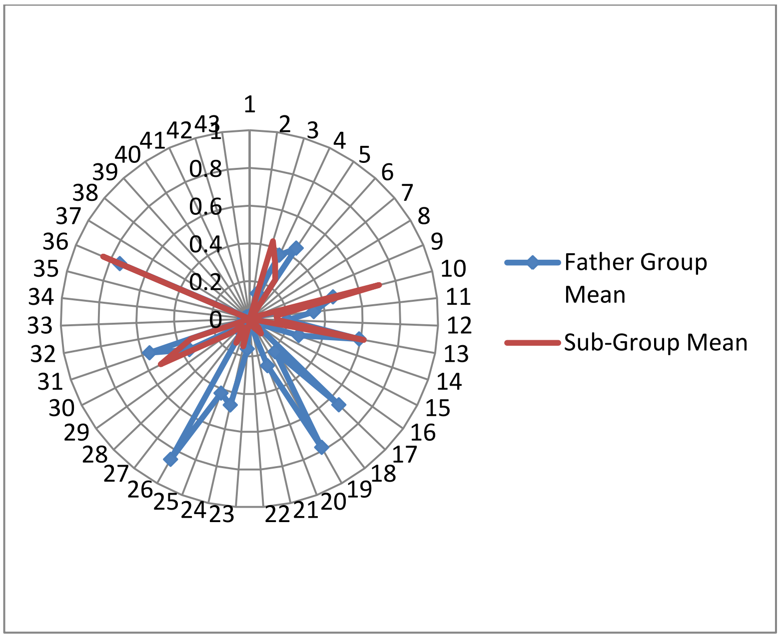
V19	Percentage of house with heating	Housing Typology	**POS**
V26	Percentage of building with 1 or 2 roofs	Housing Typology	**POS**
V17	House owners	Housing Typology	**POS**
V24	Percentage of old houses 1919 - 1971	Housing Typology	**POS**
V25	Percentage of recent houses > 1972	Housing Typology	**POS**

Histogram Sub-group 4.3. Variables differences from 1st level Group 4.

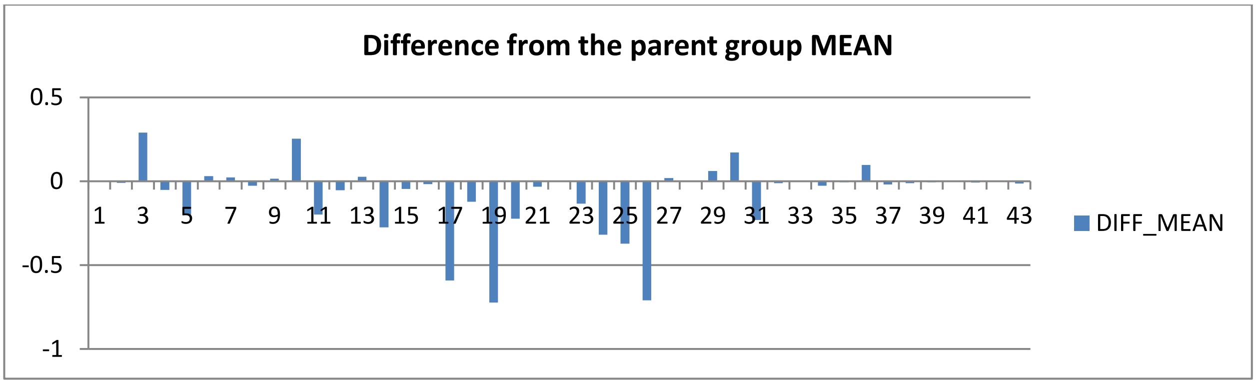
LABEL: Well-off retired families.

A2.3.12 Sub-group 4.4 sheet [sub-group 12]

The sheet describes the sub-group 4.4 (s-gr.12) main features as obtained from the variables measures.



Radial plot Group 4.4.



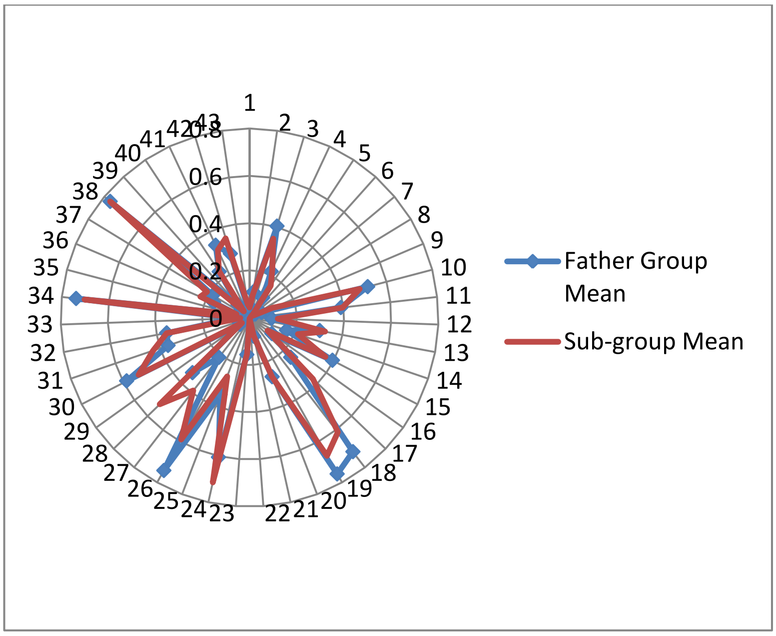
V3	Percentage of resident population aged 25-44 years	Demographics	POS
V5	Percentage of resident population aged > 65 years	Demographics	NEG
V10	Unmarried	Household composition	POS
V14	Couple household	Household composition	NEG
V17	House owners	Housing Typology	NEG
V19	Percentage of house with heating	Housing Typology	NEG
V20	Average rooms per house [normalized by range method]	Housing Typology	NEG
V24	Percentage of old houses 1919 - 1971	Housing Typology	NEG
V25	Percentage of recent houses > 1972	Housing Typology	NEG
V26	Percentage of building with 1 or 2 roofs	Housing Typology	NEG
V31	Percentage of people with Low or No education level	Socio economic	NEG
V30	Percentage of people with Medium education level	Socio economic	POS
V11	Married	Household composition	NEG

Histogram Sub-group 4.4. Variables differences from 1st level Group 4.

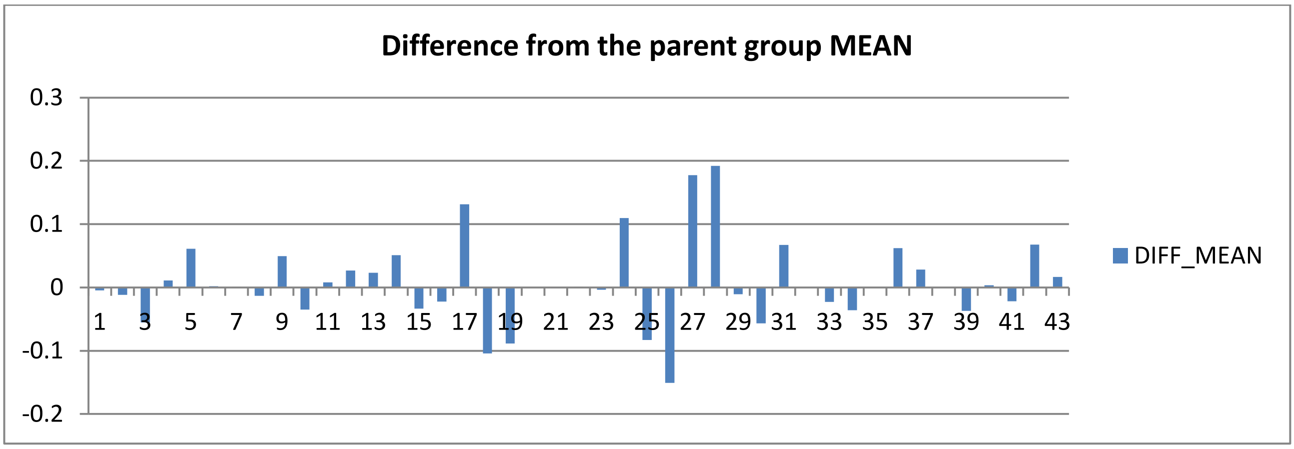
LABEL: On-a-budget couples.

A2.3.13 Sub-group 5.1 sheet [sub-group 13]

The sheet describes the sub-group 5.1 (s-gr.13) main features as obtained from the variables measures.



Radial plot Group 5.1.



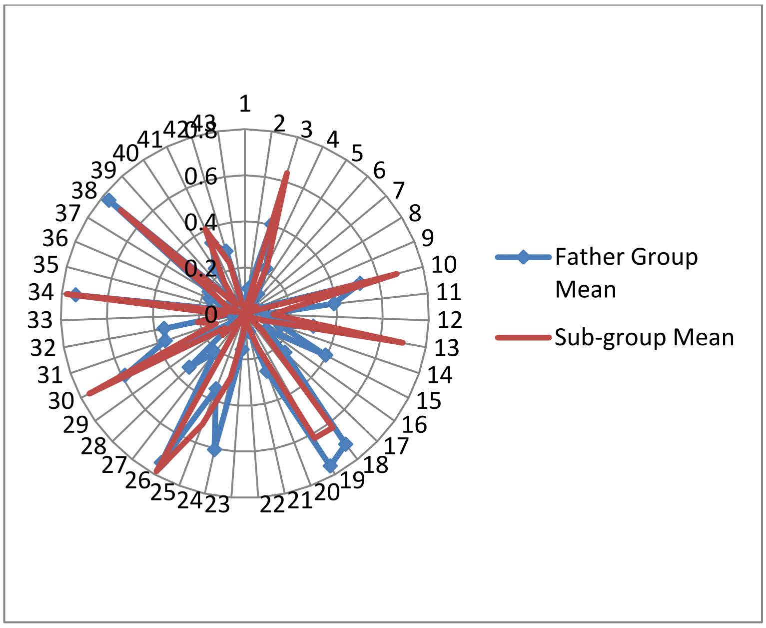
V26	Percentage of building with 1 or 2 roofs	Housing Typology	**NEG**
V27	Percentage of condominium	Housing Typology	**POS**
V28	Percentage of building with > 3 apartment numbers	Housing Typology	**POS**
V17	House owners	Housing Typology	**POS**
V18	House tenants	Housing Typology	**NEG**
V34	Percentage of employed	Socio economic	**NEG**

Histogram Sub-group 5.1. Variables differences from 1st level Group 5.

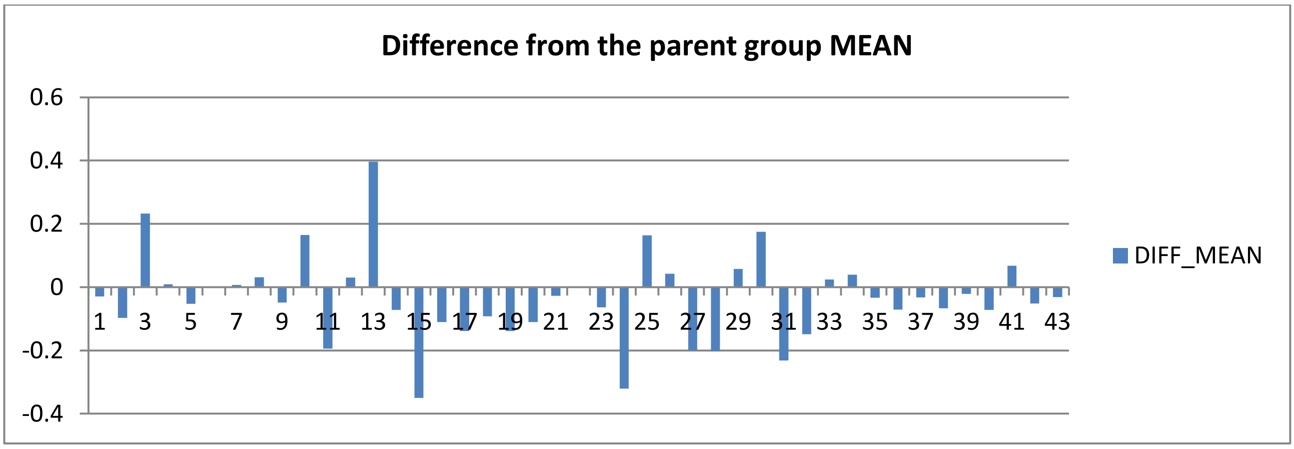
LABEL: Middle-class singles.

A2.3.14 Sub-group 5.2 sheet [sub-group 14]

The sheet describes the sub-group 5.2 (s-gr.14) main features as obtained from the variables measures.



Radial plot Group 5.2.



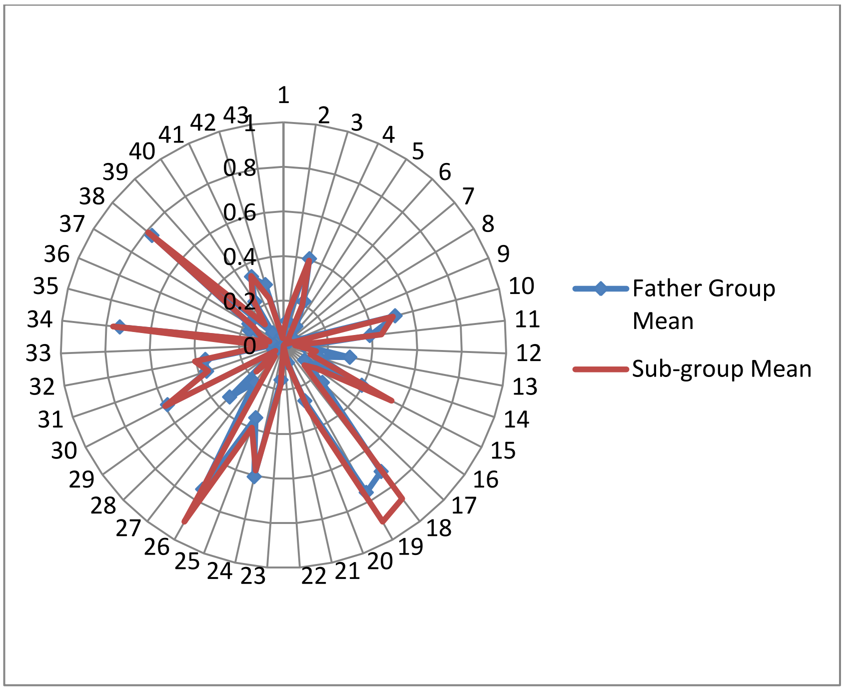
V3	Percentage of resident population aged 25-44 years	Demographics	**POS**
V13	Single person household	Household composition	**POS**
V15	Household of 3-4 persons	Household composition	**NEG**
V24	Percentage of old houses 1919 - 1971	Housing Typology	**NEG**
V28	Percentage of building with > 3 apartment numbers	Housing Typology	**NEG**
V31	Percentage of people with Low or No education level	Socio economic	**NEG**

Histogram Sub-group 5.2. Variables differences from 1st level Group 5.

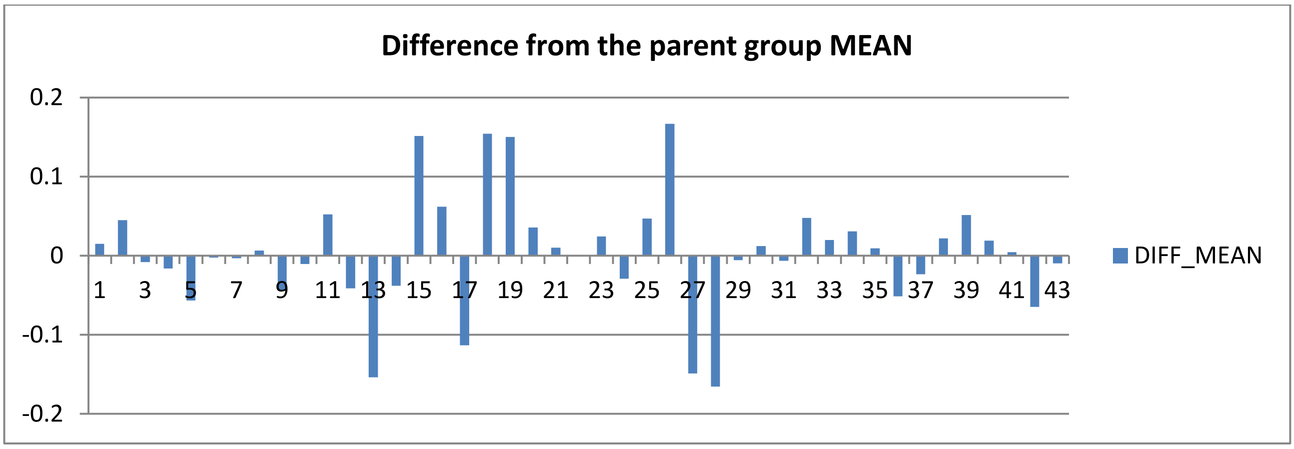
LABEL: Single white-collars.

A2.3.15 Sub-group 5.3 sheet [sub-group 15]

The sheet describes the sub-group 5.3 (s-gr.15) main features as obtained from the variables measures.



Radial plot Group 5.3.



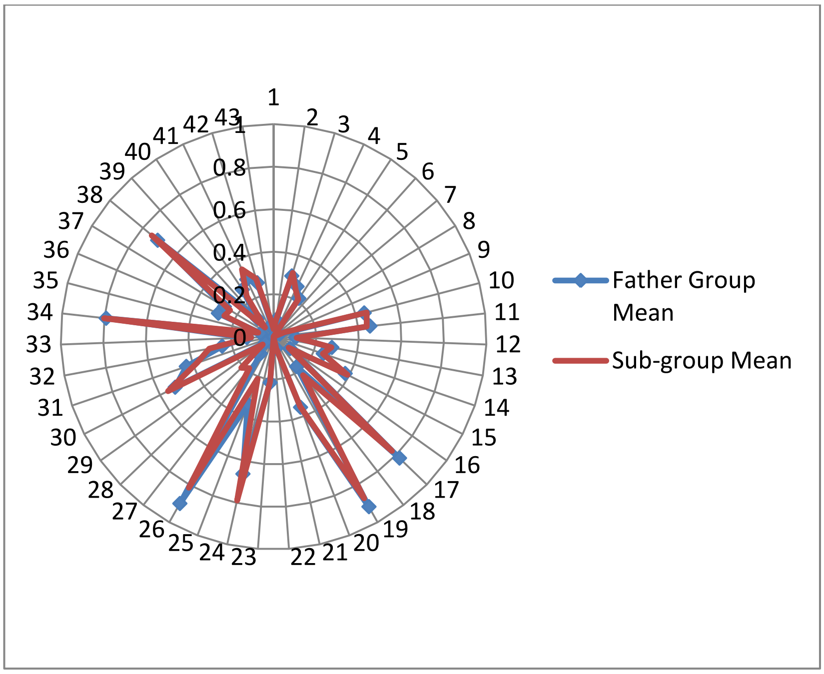
V13	Single person household	Household composition	**NEG**
V15	Household of 3-4 persons	Household composition	**POS**
V18	House tenants	Housing Typology	**POS**
V26	Percentage of building with 1 or 2 roofs	Housing Typology	**POS**
V28	Percentage of building with > 3 apartment numbers	Housing Typology	**NEG**
V17	House owners	Housing Typology	**NEG**
V19	Percentage of house with heating	Housing Typology	**POS**
V27	Percentage of condominium	Housing Typology	**NEG**

Histogram Sub-group 5.3. Variables differences from 1st level Group 5.

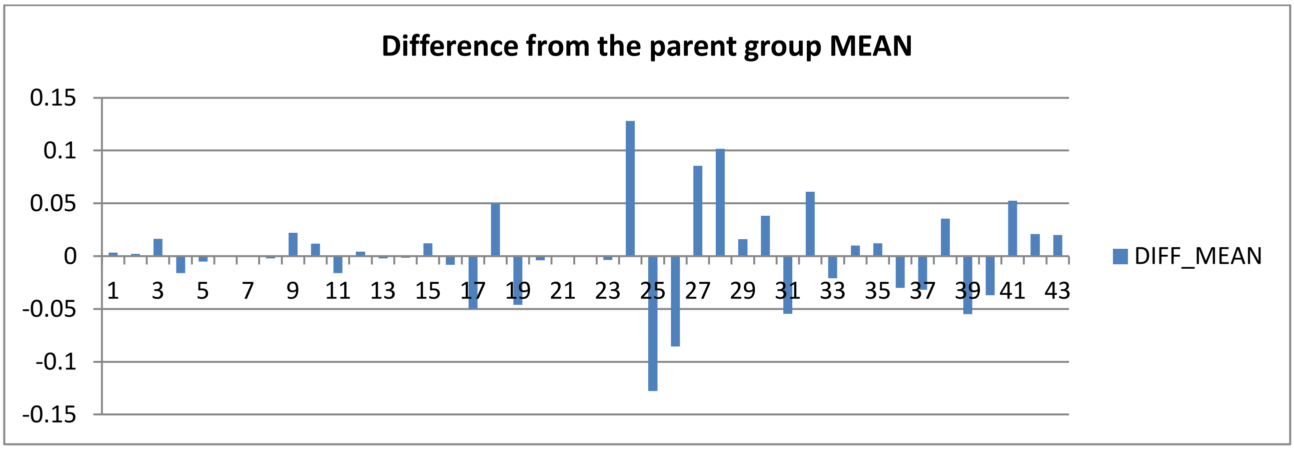
LABEL: Young well-off families.

A2.3.16 Sub-group 6.1 sheet [sub-group 16]

The sheet describes the sub-group 6.1 (s-gr.16) main features as obtained from the variables measures.



Radial plot Group 6.1.



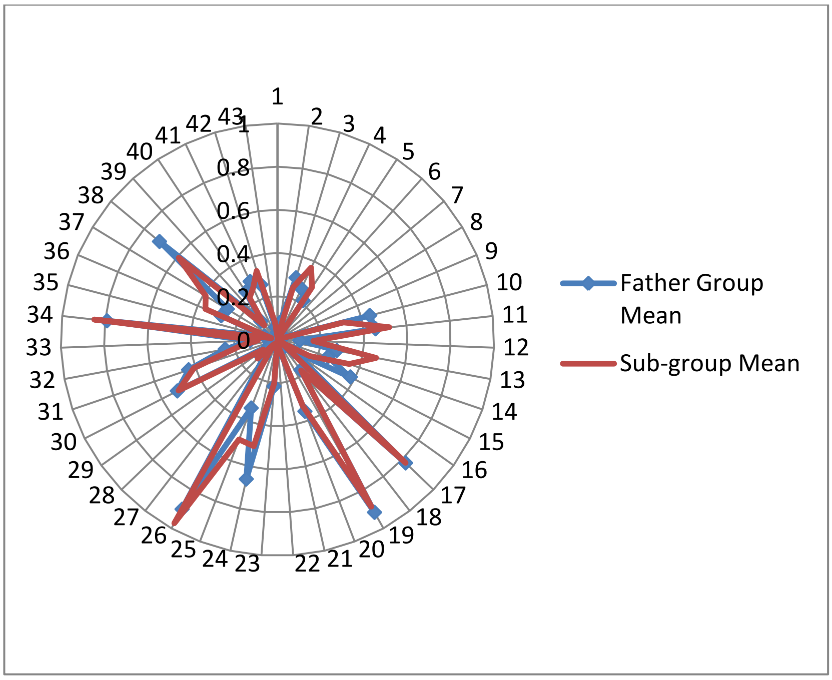
V24	Percentage of old houses 1919 - 1971	Housing Typology	**POS**
V25	Percentage of recent houses > 1972	Housing Typology	**NEG**
V28	Percentage of building with > 3 apartment numbers	Housing Typology	**POS**
V26	Percentage of building with 1 or 2 roofs	Housing Typology	**NEG**
V27	Percentage of condominium	Housing Typology	**POS**

Histogram Sub-group 6.1. Variables differences from 1st level Group 6.

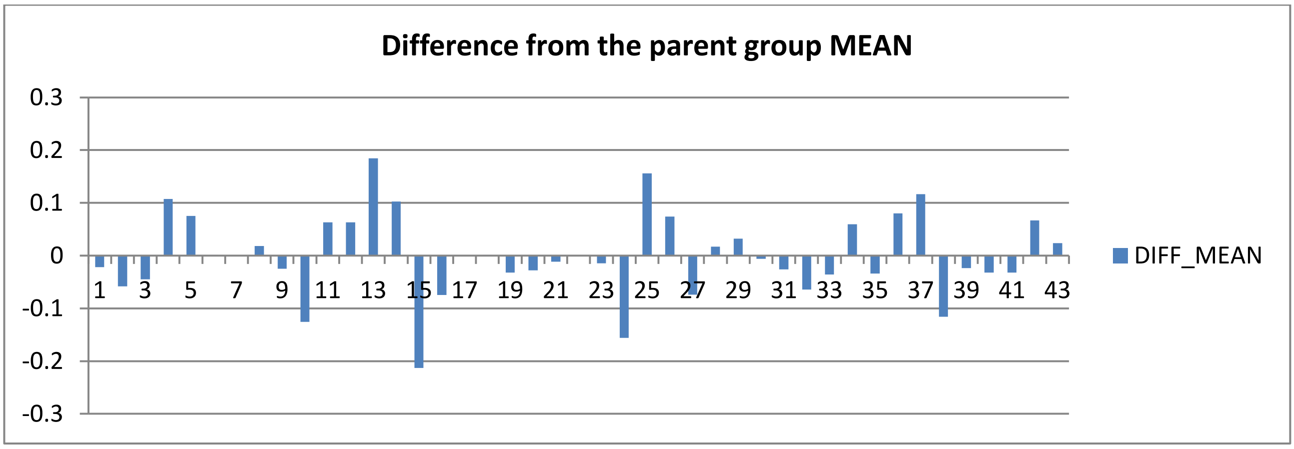
LABEL: Elder working couples.

A2.3.17 Sub-group 6.2 sheet [sub-group 17]

The sheet describes the sub-group 6.2 (s-gr.17) main features as obtained from the variables measures.



Radial plot Group 6.2.



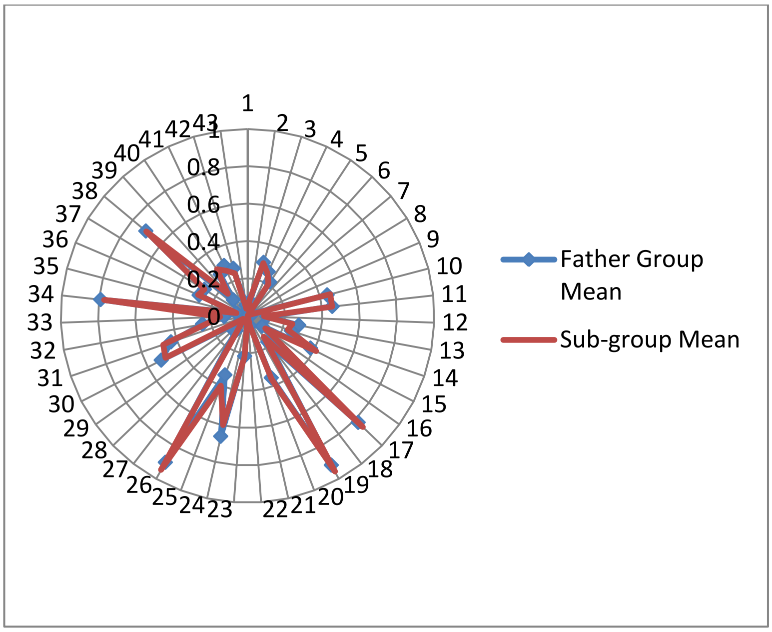
V15	Household of 3-4 persons	Household composition	**NEG**
V13	Single person household	Household composition	**POS**
V24	Percentage of old houses 1919 - 1971	Housing Typology	**NEG**
V25	Percentage of recent houses > 1972	Housing Typology	**POS**
V14	Couple household	Household composition	**POS**
V4	Percentage of resident population aged 45-64 years	Demographics	**POS**
V10	Unmarried	Household composition	**NEG**
V37	Percentage of businessman of freelancer	Employment condition	**POS**
V38	Percentage of salaried worker	Employment condition	**NEG**

Histogram Sub-group 6.2. Variables differences from 1st level Group 6.

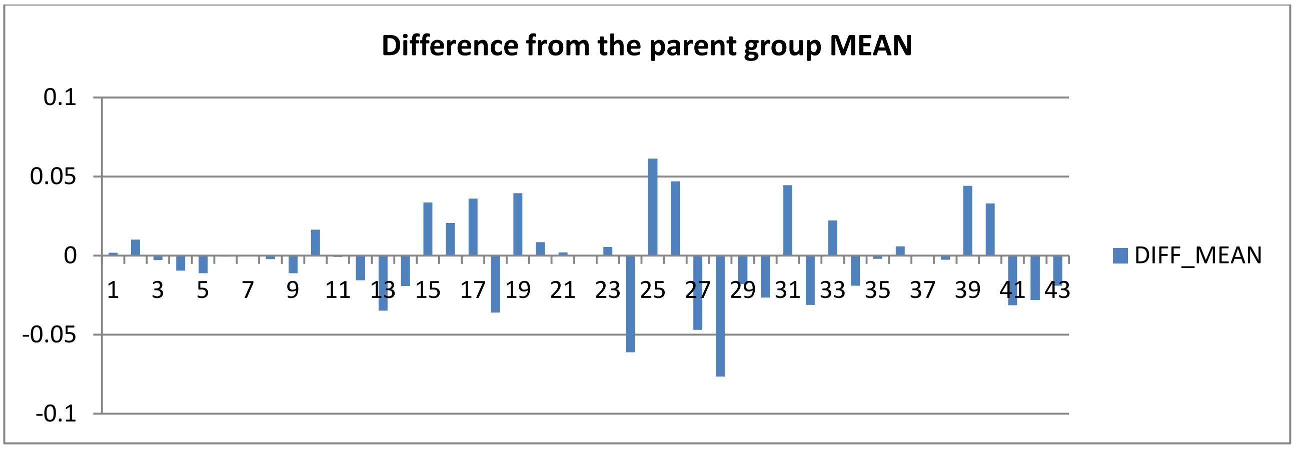
LABEL: Tradesman retired families.

A2.3.18 Sub-group 6.3 sheet [sub-group 18]

The sheet describes the sub-group 6.3 (s-gr.18) main features as obtained from the variables measures.



Radial plot Group 6.3



V28	Percentage of building with > 3 apartment numbers	Housing Typology	**NEG**
V24	Percentage of old houses 1919 - 1971	Housing Typology	**NEG**
V25	Percentage of recent houses > 1972	Housing Typology	**POS**
V26	Percentage of building with 1 or 2 roofs	Housing Typology	**POS**
V31	Percentage of people with Low or No education level	Socio economic	**POS**
V39	Percentage of employers in Agricolture	Employment condition	**POS**

Histogram Sub-group 6.3. Variables differences from 1st level Group 6.

LABEL: Blue-collars retired families.