Fuori Luogo Rivista di Sociologia del Territorio, Turismo, Tecnologia
Volume 9 – numero 1/2001

ISSN: (Print) 2532-750X (Online) 2723-9608
Journal Page: http://www.serena.unina.it/index.php/fuoriluogo/

Understandig Territory from an Online Perspective: Twitter and New Ways to do Research on Urban Phenomena

Crescentini, N.[a]; De Falco, C.[b], Ferracci, M.[c]

(a) Dottoranda al Dip.to di Scienze Sociali dell'Univ. degli Studi di Napoli Federico II, mail noemi.crescentini@unina.it
(b) Assegnista di ricerca in Sociologia presso il Dip.to di Scienze sociali dell'Univ. di Napoli Federico II, mail ciroclemente.defalco@unina.it
(c) Dottorando al Dip.to di Scienze Sociali dell'Univ. degli Studi di Napoli Federico II, mail marco.ferracci@unina.it

To cite this article: Crescentini, N., De Falco, C. C. (2021). Understandig Territory from an Online Perspective: Twitter and New Ways to do Research on Urban Phenomena, *Fuori Luogo Rivista di Sociologia del Territorio, Turismo, Tecnologia*. Volume 9 – Issue 1/2021. Pages 129 - 147. DOI: 10.6093/2723-9608/7691

To link to this article: https://doi.org/10.6093/2723-9608/7691

Manuscript accepted:29/1/2021
Manuscript revised: 12/6/2020
Published: 26/06/2021

ABSTRACT

In the era of the data revolution (Kitchin, 2014), new kinds of data and data sources allow researchers to share innovative ways of studying society and its dynamics to comprehend the peculiarities related to the territory. In fact, the analysis of space-time data extracted from social media can support the comprehension of complex social phenomena such as, for example, mobility (Noulas, Scellato, Lathia, & Mascolo, 2012; Yuan et al., 2017; Zhang et al., 2016) or where and when certain social activities are carried out, such e.g., the participation in big events, tourist experiences or the monitoring and management of emergencies and security. Following the actual dissemination of use of social media in the cities, the analysis of the spatial dimension is considered particularly promising in the field of urban analysis (Singleton et al., 2018) and in the emerging line of digital studies. Nowadays, the intentional action of many people is increasingly contributing to the production and sharing of spatial data with a relevant impact on the construction of Territorial Information Systems (Zupi, 2017). This new category of spatial data is commonly known "Volunteered Geographic Information" (VGI) (Goodchild, 2007). The analysis of this kind of data leads to an innovative scenario for the collection and diffusion of geographic information from millions of users all over the world. It provides valuable insight into their perceptions and needs, opinions on places, events, and daily routes, which allow researchers to better follow traces about the knowledge of the local identities of a given place (Campaign, 2014), information flows, and social networks within society (Stefanidis et al., 2013). However, a clear awareness of the limits of these tools of investigation is needed. For instance, problems may arise regarding representativeness of population, or even technical issues, such as for scarping systems which are not able to offer all the effective amount of data available to the researcher. Despite this, these kinds of studies are taking hold in the social sciences which are paying particular attention to both data and source contextualisation.

KEYWORDS

Sociological Space,
Algorithms,
Geo-tagged Twitter

Noemi Crescentini, Ciro Clemente De Falco, Marco Ferracci[1]

Understandig Territory from an Online Perspective: Twitter and New Ways to do Research on Urban Phenomena[2]

1. Space and society: old and new way to study this relationship

In sociological theory the role of space - understood in a physical sense and relative to the phenomenon under consideration - has never been pre-eminent or in any case well defined. Despite the recent elaborations of authors such as Giddens (1979, 1981), Harvey (1978,1993), Gieryn (2000) and Jacobs (1961) who highlighted the centrality of the spatial dimension in the study of social phenomena, this relevance has not yet been fully recognized. Consider, for example, the line of studies that refers to authors such as Sorokin, Berger and Luckman according to whom spatial determinants are, in sociological analysis, of secondary importance compared, for example, to temporal and functional ones (Strassoldo, 1990). Similarly, Parson, in his early elaborations on the social system argued that space was an irrelevant aspect of social action (Strassoldo,1983). Popper on his part promotes the idea of an individual completely isolated from society whose existence he denies and therefore denying any possible influence (Mauceri, 2014). This absence of space in the history of social theory and sociology itself in the twentieth century has led some authors to speak of an a-spatial sociology (Strassoldo, 1983; Mela, 2006). Yet numerous studies have highlighted how the social and non-social characteristics of physical space influence and shape individual action together with the other types of "space" that make up "social space" and that permeate human experience[3]. According to Strassoldo, "social space" is made up of six iterative layers: a) ethological or biological space b) personal space c) lived or existential spaces d) symbolic or cultural space e) ecological space f) political and organizational space. The physical /ecological space has been the object of attention of some of the classical authors of sociology, including Simmel, Durkheim and Park. In particular, Park, (1970) even if with veins of biological determinism (Helmes-Hayes, 1987), conceives space as the main dimension of the work on human ecology carried out by the Chicago school.

Durkheim, on the other hand, makes it the center of his research program – which he never actually started - on social morphology. For Durkheim the social morphology is constituted "by the number and nature of the elementary parts of which society is composed, how they are arranged, the degree of connection they have reached, the distribution of the population over the surface of the territory, the number and nature of communication routes, the shape of the houses and so on "(Durkheim, 1893, 31-32). Of Durkheimian derivation and in particular his thesis on social and physical "density" as the cause of the specialization, competition and division of labor, are the topics of studies that have placed

[1] Noemi Crescentini, dottoranda al Dipartimento di Scienze Sociali dell'Università degli Studi di Napoli Federico II, E-mail noemi.crescentini@unina.it, Ciro Clemente De Falco, assegnista di ricerca in Sociologia presso il Dipartimento di Scienze sociali dell'Università di Napoli Federico II ; E-mail: ciroclemente.defalco@unina.it, Marco Ferracci, dottorando al Dipartimento di Scienze Sociali dell'Università degli Studi di Napoli Federico II, E-mail marco.ferracci@unina.it.
[2] Received: 29/01/2021. Revised: 12/06/2020. Accepted: 26/06/2021. This article should be considered as collective elaboration especially for the Introduction and for the Conclusion. However, paragraphs 1, 5, 6.2 are to be attributed to Ciro Clemente De Falco, paragraphs 2, 6.1 are to be attributed to Noemi Crescentini, paragraphs 3 and 4 to Marco Ferracci and paragraph 6.3 to Noemi Crescentini and Marco Ferracci.
[3] Among these studies, for example, we certainly find those that refer to the works of Cox and Galster, as regards electoral behavior, and that of Wilson (1987), as regards the study of marginality.

under empirical investigation the influence of the characteristics of ecological space on individual action and social phenomena.

This research field embraces, in addition to sociology, contributions from disciplines such as psychology, political science, economics, etc. Concepts such as structural effects (Blau, 1960), compositional effects (Davis *et al.*, 1961), contextual effects (Lazarsfeld, 1961) and neighbo(u)rhood effects (Wilson, 1987) represent the different attempts to articulate the discourse on the relationship between space and social phenomena. This articulation is also realized in the differences between the four conceptualizations. The first difference is of a methodological nature and concerns the type of analysis unit to be considered (Robinson, 1950; Menzel 1950) and which analysis techniques to use (De Vos, 1998). The second difference concerns the operationalization of "space". In studies close to Blau's theoretical elaboration - recognized by those who deal with territorial analysis as the first author to systematize the discourse on aggregate effects on individual behavior (Pintaldi, 2006) which focuses on structural effects, the focus is placed on the socio-economic characteristics of the population occupying a given space, Blau, indeed, focuses solely on the variables that have an individual consideration.

In scientific production close to that of the so-called Neighborhood effects, whose popularity is due to Wilson's famous and seminal research on disadvantaged neighborhoods, attention falls not only on aggregate properties, but also on the global properties (Lazarsfeld, 1961) of space within the frames mentioned. The body of research that has demonstrated the causal effects of space features on social phenomena is large. To give greater depth to the results obtained through the quantitative approach were the qualitative investigations that tried to open the so-called "black box" to reveal the mechanisms that connect the characteristics of space to individual action.

In a recent analysis of the literature, Galster (2012) identified 15 types of mechanisms that can be grouped into four large families: social interactive; environmental; geographical; and institutional. This line of research briefly described, which is in part close to spatial sociology (Serino, 2017) promoted by authors such as Mela (2006), could find a new impetus thanks to the phenomenon of big data and the wide availability of new geo- referenced data. Given that part of the geo-referenced data comes from social media, the spatial dimension, initially absent in analyses focused on social media, by definition digital spaces without anchors (Menduni, 2014), is becoming the object of increasing attention also in in the context of the Internet or digital studies.

The purpose of this article is therefore twofold, the first is to analyze the way in which this new approach is being constituted with particular attention to the urban dimension. The second, connected to the first, is to identify which are the key elements to deal with for those who move in this approach. In particular, we will try to give the reader the answer to the following questions: a) on which topics do the geo-social media consent to do research? b) how to avoid epistemologically naive conceptions of space? c) what are the methodological precautions to keep in mind during the analysis and data collection phase? d) how to choose the existing statistical tools? To answer these questions, we analyze the research that has worked on geo-social media. The article will therefore be structured as follows: in the second section the way in which the spatial dimension has been affirming itself within the analysis of social media will be discussed, while in the third the attention will focus on the methodological and epistemological problems that the use of geosocial media pose. In the fourth section we will see the role played by the urban dimension in this new analytical scenario while in the fifth section we will discuss the methodology applied to answer the research questions. We used a PRISMA systematic review in an innovative way to create our dataset of articles. Based on our objectives we will not show a theoretical review in this article, but we want to offer some methodological tools to manage empirical research, tools that will be explored in section six

2. Managing research thanks to social media

The relationship between space and social relations, as we saw in the previous section, has always been the subject of discussion for the social sciences and space is now becoming the subject of increasing attention in relation to the Internet and digital studies. In recent years, in fact, the studies have developed using available data in the digital scenario from the web and, specifically, from social media. They are considered not only as an object of study but also a useful source to comprehend social phenomena (Rogers, 2016). In this digital scenario, social media tends to create new channels of communication aiming to collect, generate, share, and circulate specific information, including information relating to space. In recent years, researchers have used social media data as a source to study different types of human activities and social phenomena, becoming a significant resource in order to improve specific circumstances such as in natural disasters. Among these studies, there is indeed a new strand of social research that has chosen to study space and its urban dynamics moving in the field of social media and using the spatial-temporal Big Data. The spatial-temporal Big Data consists of elements such as volume, variety, speed, veracity and value, and today the discovery of knowledge from spatial-temporal big data focuses primarily on summaries, blurred outlines, rare associations and predictions, which are expansions of traditional space-time data mining (Shi Y., Deng M., *et al*., 2016).

The current media context, shaped mainly by social media, affects interplay practices in terms of framing due to the content sharing and different use of digital instruments of a shared repertory based on complex networks architectures. In this way, according to Rogers (2009) the large amounts of content expressed by social media opened has led to an attitude that *following the medium* allows researchers to comprehend social change through internet crossing the online/offline environments. Furthermore, these environments are no longer separate but interact with each other. Following Natale and Airoldi (2017: 11-18) social researchers today are particularly interested in what aspects of social phenomena can be focussed on in an investigation through the analysis of digital data on the basis of several potential dimensions by which to develop any empirical investigation.

Rather than regarding social media data analysis as a surrogate for traditional methods of inquiry, therefore, the experience of digital social research to date suggests the need for an account of how it might augment, and be augmented by, traditional social research methods because what is really distinctive about social media data is the production of naturally occurring or 'user-generated' data at the level of populations in real or near-real-time (Adam Edwards *et al*., 2013). In this way, social research on digital media, such as social media, has the opportunity to investigate social processes at the very moment when they occur (or later). Among the latter, in the background of our study there are the so-called geospatial data, which can be extracted from social media like Facebook, Twitter or Instagram and Foursquare through specific scraping tools and procedures. We can now provide an overview of dynamic models of urban environments and life with higher spatial and temporal resolutions than conventional data sources (such as census data and field research) (Batty, 2013) which allows a more complex analysis of the composition of the data, also formed by various languages and contents such as videos, animations, images and other creative elements that produce systematic meanings. Several different geo-tagged contents related to social media, also called geo-social media, contain both descriptive comments and information about the position and therefore, «can support the understanding of the needs of users, public opinion or where it is necessary to develop any resolutions. However, differently from polls and citizen interviews, social media data are forms of direct communication, therefore, these data are unstructured, vary in quality and are often of unidentified relevance for local government needs» (Zhang, Feick, 2016). The integration of GIS functionality in social media provides the geographical location of the content added by users. As a result, published posts about facts, events, opinions, or multimedia content may be related to the spatial feature. In this way it has been possible to move from the dimension of the virtual space of the Internet to the real one of the users, integrating the places to everyday life (Massa *et al*., 2014).

The geo-localized posts described above are called Social Media Geographic Information (SMGI) (Campagna, 2014). In particular, two types of data integrated with traditional and pre-existing information regarding social analysis and urban space have been create: Volunteered Geographic

Information (Goodchild, 2007) and Geographic Information deriving from Social Media (Campagna *et al*., 2016). Some social media such as Instagram, Flickr or Facebook provide the function of adding geographical information to an audio-visual content only if directly carried out by users. This geographical position may also differ from the place where the user is actually located.

Twitter is more reliable when it comes to adding geographical information. This allows anyone with an appropriate device to generate a post of 140 characters or less and to share it with other users and allows the specification of the position outlined by latitude and longitude. The user has both the opportunity of self-tagging in a given place and to automatically generate the location of the content if activated GPS functions on the device. «This location, known as a geotag, allows further analysis to be performed relating the content of the posts, known as tweets, to specific areas of the city where they were generated» (Meyer, Balaguè *et al*., 2018). Nowadays most intelligent mobile phones are GPS-enabled, so geographical location information is often included as an additional tag in tweets. Combined with the time annotation, this type of spatial-temporal information can be embedded in tweets to describe where and when the tweets are broadcast. So, Twitter data has become a kind of spatial-temporal Big Data (Shi Y., Deng M., *et al.,* 2016). Ways to extract hidden, unknown and significant events from the huge mass of Twitter data has thus become a research hotspot in computer science, human science and GIS in recent years. Given the widespread use of social media in cities, the analysis of social media activity is considered particularly promising in the emerging field of urban analysis (Singleton *et al*.2018). Since 2010, Twitter has been distinguished both qualitatively and quantitatively among the various social platforms as one of the most widely used sources for urban and mobility studies. Geo-localized tweets are useful tools for urban studies allowing space-time tracking / tracing of users (e.g., the times and places that concentrate the greatest number of users) (Condeço-Melhorado *et al*., 2020).

3. Being aware about data and algorithms

Nevertheless, working with big data shows particular dimensions that should be considered. Big data and algorithms offer a high degree of opacity (Burrell, 2016) meaning a high degree of cultural and social features that influences their processes which were hidden by their mathematical, mechanical, aseptic and data driven foundations (Dourish, 2016). We will not discuss these topics in depth in this article but a mention is needed, in particular in regards to social platforms, or as in this case Twitter. First, Big data could lead to a data driven conception of interpreting our world. Claiming "data speaks for itself" entails conceptual construction, effective values that data contain and the ways by which they were collected -for instance, the position and the construction of a particular device- were omitted. Secondly, one of the biggest problems regards the amount of big data that could be effectively extracted thanks to scraping and APIs systems. Most extraction systems have a time range -for instance 7 days- to extract data, in some case this problem could be overcome by paying a lot of money, but in the other cases these data cannot be retrieved. This problem sheds light on the need to consider these issues when arranging research. However, even if some tweets were to be lost these systems allow the researcher to work with a large and important number of data, making these reliable. In addition, sometimes extraction systems are not able to collect all the data available in a defined period of time, for various, even very technical reasons. However, the battle between different values and points of view about how algorithms make work possible are present all over. In a more pragmatic way, data would often need to be cleaned and harmonized depending on the source.

The third problem regards the data itself in particular. Data from social networks are not representative of the whole population (Malik *et. al*., 2015; Jiang *et. al*., 2019; Zhang *et. al*., 2018) because many people might not be interested in using social media, and, or in having a smart device, such as a smartphone, tablet, computer, or they do not have the skills to use these. As far as geolocated data are concerned, researchers who use these as the main source of their investigation have to be aware that the position of tweets could be different from the real position where the tweets were posted. Moreover, only 1% of tweets have GPS localization.

The last and most obvious issue regards the topics that might be found in the social media environment and which change from platform to platform. For instance, as claimed above, Twitter has an important potential to study some phenomena of the urban dimension, while it is less effective in studying, for instance, some kinds of age groups. Methodological awareness refers to the capacity of scholars to include these kinds of problems in their studies with Twitter data. All these features have to be take into consideration to produce results and manage inferences.
Tweets can be collected thanks to different strategies: the first is web scraping and the second is application programming interface (APIs). These two systems offer two automated techniques to gather data, while we will overlook the manual extraction. Web scraping analyzes websites to collect data; this technique entails an analysis of HTML pages using the tag on tweets. In technical terms, a robot or a script looks for the topic of interest by interacting with the different pages and selecting the suitable content (Dongo *et al* 2020). This tool might be particularly useful when the contents belong to a particular page allowing the extraction of a large amount of information.
The second strategy is API. This involves two parts: the first regards the programming interface of the application, in this case Twitter, and the second is external software capable of using these keys to collect data, such as rtweet (*ibidem*). Twitter in particular offers different types of access to their data with different costs. The next step is to decide how to identify the location to extract tweets, and three methods may be used for this: geotagging, geoparsing and geocoding. In using these tools methodological awareness is needed as claimed above, as each of these tools bring their own specific problems and advantages. The first, geotagging, regards the decision of users to tag their location thanks to GPS of their device when posting a tweet (Bakerman *et al.,* 2018). In spite of the fact that this kind of position reference is the most reliable, only 1% of tweets were geolocated and Twitter algorithms of localization could make the position incorrect, This kind of data extraction also has 3 modalities to gather data in a particular location: bounding box, to define an area within which extract data; using keywords, or rather words that the author has chosen as interesting for their work, searching for these within a range of tweets with GPS location; finally from secondary data, reusing data collected from other studies. Geoparsing refers to the possibility of using free text of twitter to identify the posting location. It allows the identification of toponyms: "toponym is any named entity that labels a particular location" (Gritta *et al* 2020: 690). Finally, Geocoding consists of transforming a defined textual representation of an address into a valid spatial representation (Middleton *et al* 2018). It consists of matching a word that describe a location with GPS data (Zhang, Gelernter 2014)

4. New spaces in older cities

Since the beginning of sociological studies urban dimensions have been investigated from different points of view (Weber, 1914; Simmel, 1903; Durkheim, 1893; Jacobs, 1961; Lutters, Ackerman, 1996). Urban spaces have been characterized by their own structural and physical dimension within which many social activities take place. Cities, for instance, show characteristic environmental dimensions, and histories tied to their "political space" shaping different arrangements (Bagnasco, Les Galès, 2002; Sassen, 1994; Kazepov, 2005; Vicari Haddock, 2013). However, nowadays, spaces that contribute to studying the urban dimension are becoming vaster and more complex. Smart cities, big and open data societies, are taking over the urban space making this the most sociological of all times (Taylor, Richter, 2015). This offers important ways to govern, study and define urban dynamics taking into consideration the chance to study phenomena in the moment when they occur. At the beginning of the 2000s, Social GIS applications for the analysis of actions and habits in the urban environment were the first applications developed to focus on both territory and digital activities (Zupi, 2017). Although, in a more recent time, cities have been wearing a *new digital skin* (Rabari, Storper, 2015) this changes not only the methods to understand how the city "lives", but, also, how the space are made up. Indeed, this kind of Big Brother viewpoint that observes each movement, probably evokes the images of Orwell, but, in practice, this new skin has a fundamental role in defining space.

Digital and real space are becoming more closely related to each other, so as to become an assemblage without which cites might lose their own identities. However, the “social production of space” thanks to the digital dimension should be divided according to their different ways of production. Social GIS space represents a map thanks to people who are able to identify geographic information, but these kinds of maps could contain, for instance, information about some kind of residential information, or socio-demographic features of space of the different kinds of zone where a particular political-social intervention takes place. New forms of data production have led to the conception of cities as *data factories* (Economist, 2012) calling the attention to the relationship between smart objects and human actions in defining the space of the city.

One of the most used and interesting applications is the case of traffic (Courmont, 2021) a great deal of different data from different sensors, devices and video cameras defines the streets to identify and govern these phenomena in real time thanks to the big and open data that the city produces. However, in addition to this, in the last decade, the decade of *platform society* (Van Dijck *et al*., 2018), cities have been invested with new kinds of data and space production. Thanks to smartphones each person can indicate their position, information about their feelings, ideas, information regarding place all of this, voluntarily defining a mixed communicative and physical space (Meyer *et al*., 2019). It offers an important opportunity to study the urban dynamics and the conception of space itself. From a social network point of view, the spaces of the cities can be more closely tied to traditional and real spaces, neighborhoods, streets and in general all physical spaces (Lansley, Longley, 2016; De Albuquerque *et al*., 2015; Fu *et al*., 2018), although they might also consist of events, small meetings or unexpected events (Zhou, Xu, 2017). Big data from social networks are able to uncover and make new spaces that are otherwise unreachable visible. For instance, tourist mobility within and between cities can be studied thanks to user generated data, in particular tweets (Fei *et al*., 2018).as can, the activism of individuals in relation to some specific spaces of the city (Haffner, 2018). Sentiment analysis of tweets might help to understand the perceived quality of life of different zones of cities (Guhathakurta *et al*., 2019).

This is the aim of this paper, we will show how a particular social platform, Twitter, has been used by scholars to study the urban dimension, taking into account the advantages and the limits of carrying out research in this way.

5. Methodology

In order to investigate the main features of this emerging line of studies a systematic literature review based on content analysis was performed. So the review presents two main objectives: 1) to analyze central points such as the topics investigated; emerging concepts of space; awareness of the limits of geo-localized data and spatial statistics techniques that were used, and 2) to provide some coordinates to follow for setting up searches with geo-data. The systematic literature review was carried out between November 2020 to February 2021 by three researchers from the University of Naples “Federico II” - according to PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analysis) methodology. Prisma is a methodology for systematic reviews that ensures a strict review plan, offering methodological accuracy, transparency and the possibility to easily replicate the research (Tranfield *et al*., 2003). There are four steps that make up the PRISMA method: identification, screening, eligibility, and inclusion (as described below and in figure 1. During the first step, known as the identification phase, to properly cover the issue the Scopus database was used because it provides an excellent coverage for social science research (Burnham, 2006). To collect the articles built on geolocated tweets, the choice of keywords fell on “geo”; “twitter” and “spatial”. Thus, the following search string was defined: ((“geo” or “spatial”) AND (“twitter”)). The search produced 1,589 articles. To complete this step, after entering the search strings, inclusion and exclusion criteria were set to obtain the relevant literature in the chosen databases.

Table 1. The process of data collection

Database	*Keywords*	*Records*
Scopus	TITLE-ABS-KEY (“geo” OR “spatial”) AND (“twitter”)	1,589

Legend: ABS= Abstract; KEY = Keywords

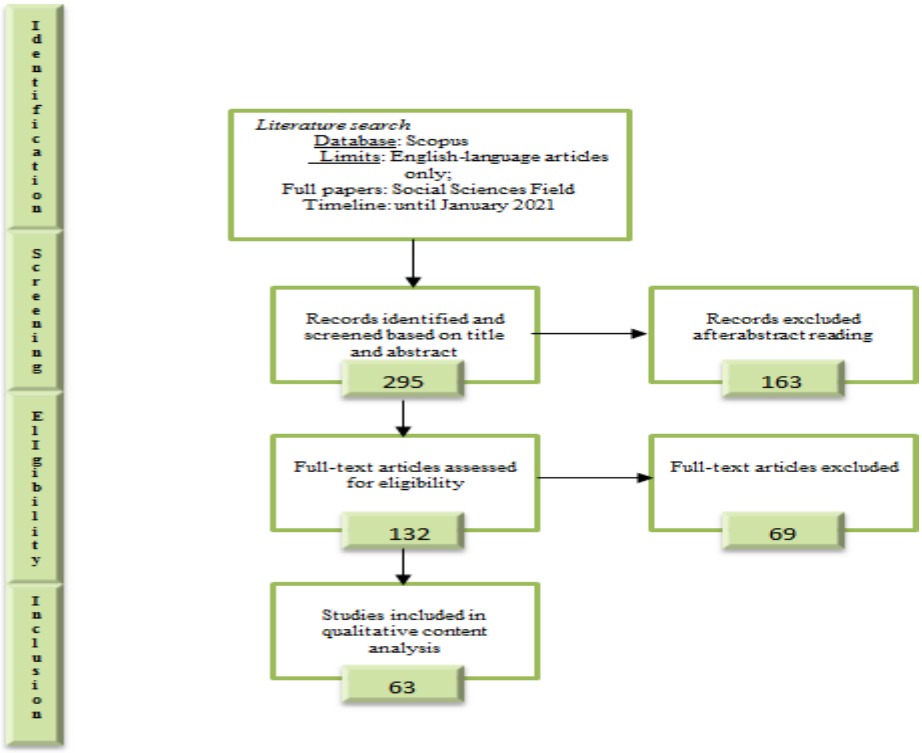
In light of the research objectives, only studies in the “social sciences” field were selected, and for scientific publications, only full papers were considered. The extraction took place in October 2020 and given the recent nature of this approach; no time limit was set for the research. In other words, we collected all the articles on the subject written up to the time of the study. Given the scarcity of results in the Italian language on Scopus and on the other international search engines like Emerald or Web of Science or national collections like Torrossa or RivisteWeb, it was decided to analyze only scientific products written in English. After this step 1,303 article were deleted. On this basis, 295 records were screened and after the reading of the abstract, 163 papers were deleted because of their low pertinence to research aims and review objectives. Thus, after the reading of the full texts of all 132 remaining articles, 63 publications were included in the review process because they contribute to answering the review questions. In line with the recommendation of Robinson and Lowe (2015) the number of papers included seems to be widely acceptable.

Table 2. Inclusion criteria

Criteria	*Description*
Document type	Peer-reviewed journal articles
Language	English
Research Domain	Social Science
Timeline	Until October 2020

To answer our research questions, it was decided to use the content analysis approach and in particular the third type of content analysis. Content analysis covers almost all techniques used to extract secondary meaning from textual information. In particular, content analysis is a research technique for making replicable and valid inferences from texts (or other meaningful matter) to the contexts of their use» (Krippendorff, 2018). Content analysis has also been defined as a systematic, replicable technique for compressing many words of text into fewer content categories based on explicit rules of coding (Berelson, 1952, Weber, 1990). In line with this definition of content analysis we find content analysis of the third type, also known as “inquiry”, which is a procedure that the researcher can apply “to any type of message - verbal and non-verbal - using a semi-standardized or standardized analysis sheet to record the states in which certain properties in a set of suitably selected units of analysis occur”. In the analysis of the content of the third type, linguistic or extra-linguistic decompositions are not carried out and the classification units coincide with the context unit.

Figure 1. The assessment and selection of contributions: PRISMA flow diagram



In these kinds of analysis, the articles are analyzed through an analysis sheet that can be seen as the equivalent of a questionnaire for individuals. In our case the analysis sheet was divided in four sections: 1) the first contains information about the general characteristics of the article (year of publication, number of citations, authors, affiliation of the authors etc.); 2) the second section contains the abstract and keywords; 3) the third section contains three useful variables for understanding the epistemological orientation of the article and necessary for the comprehension of what kind of relationship exists between the space, the phenomenon investigated through the tweets and potentially any other phenomena. 4) the fourth section contains variables relating to the methodological dimension: the type of analysis performed, the type of algorithm used for the extraction and analysis, and the possible critical discussion of the limits related to data validity and data representativeness. The data collected were mainly analyzed through descriptive statistics, while the analysis of the themes was conducted through a cluster analysis carried out on the keywords used as proxy for the main themes of each article.

6. Detecting a new approach

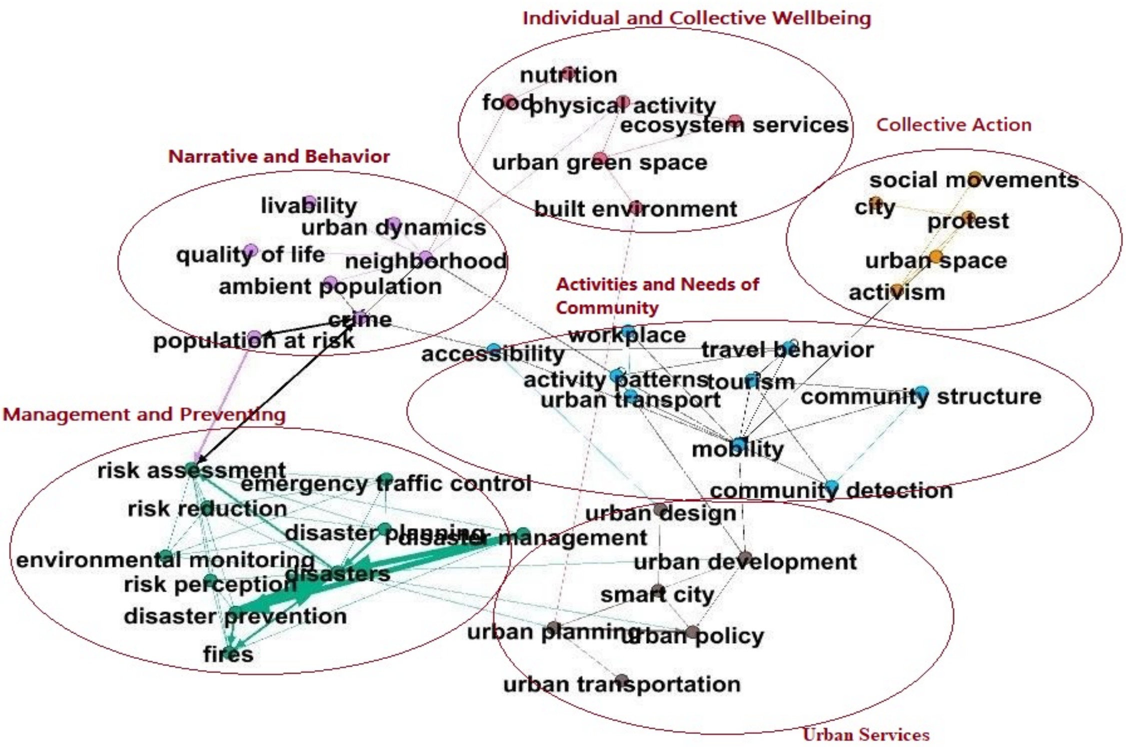
The analysis that follows focuses on some relevant aspects for setting up research that aims to exploit the potential of geo-social media and specifically Twitter. To understand what kind of phenomena can be investigated thanks to the use of these data, we will focus firstly on the themes investigated in the articles analyzed, then on the different ways of obtaining geo-localized tweets and on the awareness about the methodological limits in their use. Since the geographical dimension concerns both the tweets and the structure of the research, we will try to understand what role the characteristics of space play in the analysis of the phenomenon. Finally, an examination of the spatial statistical analysis techniques used will allow us to understand which tools to use to answer certain research questions.

5.1. *Main Topics Analyzed*

Thanks to a bibliometric analysis conducted by Vosviewer and graphically returned through the Gephi open-source software, it was possible to observe the themes that emerged from the full papers considered. This network of themes was built through the keywords present in each article considered and were subsequently grouped into 6 clusters that were named by reasoning for congruence thematic. The first cluster, called “Individual and Collective Wellbeing” includes articles regarding food, nutrition, ecosystem services or urban green space issues. For this cluster we can offer same examples of papers. The first, the paper by Kovacs-Györi *et all.* (2018) who studied park visits for urban planning purposes using Twitter data from frequent users, and thereby defining profiles of urban parks of London and their visitors. The second example is the work of Chen X.,, Yang X. (2014) whose authors use spatial-temporally tagged information and tweets providing an ideal method for measuring the exposure to the food environment in real time. This measure, as a representative of individual food access, is associated with users’ particular dietary choices conveyed in their tweets and help the investigation of human activities related to food and health research.

The cluster called “Collective Action” includes themes such as activism, social movements and protests and therefore connects to all those collective activities that take place in one or more areas of a city or a nation and that therefore includes forms of aggregation of individuals who have security, promotion or change of a socially connoted fact imposed as objective. A case study of this study is written by Haffner, M. (2018) who undertake a place-based approach to studying the factors driving the #BlackLivesMatter movement and counter-protest content (e.g., #AllLivesMatter) through geolocated tweets in Louisiana and Texas Census Designated and Incorporated Places (i.e. cities).The third Cluster called “Activities and Needs of the Community” includes articles dealing with services, commonly understood as intangible resources, with which specific needs are satisfied. For example, we find articles on urban transport, mobility, accessibility or travel behavior. For example, in the paper of Li Y., *et all.* (2017), the authors analyzed the spatial and temporal patterns of geo-tagged tweets from Midwestern college cities/towns to infer the human mobility patterns of Twitter users. In this case, they reached the conclusion that the spatial pattern of tweets distribution varies with the size of the city. Then we can find the “Urban Services” Cluster in which we find issues about smart cities, urban development, urban design or urban transportation.

Fig. 2 - *Main themes*



The paper by Anselin, Williams (2015) describes the employment of geo-located social media data to identify 'digital neighborhoods' – those areas in the city where social media is used more often and all with the application of spatial clustering techniques. The authors highlight the importance of city infrastructure and landmarks

The fifth cluster, on the other hand, is named "Management and Prevention". It contains articles that analyze geo-location data during natural disasters such as hurricanes or fires, but which are also aimed at monitoring and reducing certain environmental risks. In this case, we can notice in the paper by Li Z., Wang C. *et all.* (2017) how social media, in particular Twitter, can be a new data source for disaster management and flood mapping. The authors, in fact, using the 2015 South Carolina floods as the study case, introduce a novel approach to mapping the flood in near real time by leveraging Twitter data in geospatial processes using quantitative methods to better understand how this specific social media activity is related to flood phenomena.

Finally, the last cluster "Narrative and Behavior" focuses on narrations and the behaviors of individuals. Issues such as human activities, timeframes or urban dynamics and also specific areas such as neighborhoods, ambient population and livability can be found in this cluster. For this last cluster we can give the example of the paper by Guhathakurta *et all.* (2019) who presents a model to classify perceptions of various Atlanta neighborhoods based on social media with quantitative indicators.

5.2. *What kind of space?*

The common feature of all the selected articles is the geographical analysis of the phenomenon under investigation. It is interesting to understand the ways in which space and its characteristics are used to interpret the phenomenon variability and the different ways in which it manifests itself. This allows us to understand the position, in relation to the ecological tradition, of the articles in the analyzed corpus. In order to answer this question, three indicators are used in the analysis form: a) the impact of the characteristics of space on the analysis of the phenomenon studied b) which characteristics of space are taken into account c) the analysis conducted. The first indicator allows us to grasp the closeness to the tradition of ecological studies. The starting point is the assumption that the analysis of the simple geographic distribution of a phenomenon is prerequisite for an ecological analysis (see the first section) because the geographic variability of the phenomenon should be traced back to the kinds of characteristics of the space in which it occurs. From this point of view, the analysis of the way in which the spatial dimension is linked to the phenomenon analyzed allows us to distinguish three main orientations within the corpus that have been defined as: descriptive; reconstruction; explanatory (Cf. tab. 3)

Table 3 - Relation between space and phenomenon analyzed

Orientation	**n. of article**
Data Driven orientation	
Descriptive	29
Theory Driven orientation	
Explanatory	21
Reconstruction	13
Total	63

In the first orientation, the descriptive one that includes about half the articles (29/61), the analysis consists mainly of showing how the phenomenon is distributed over the territory considered without tracing the variability observed back to some characteristic of the territory. In our opinion, this approach is far away from the ecological tradition and it can be considered a data-driven approach, since it includes the geographical dimension without taking into account how space may affect the phenomena analyzed. However, it should be emphasized that these types of articles can be considered for further studies. and/or their methodology can be used to create event-detection systems[4]. In the articles classified as explanatory (21/63), there is a precise operational definition of the space characteristics, and the information collected on the ecological unit is included within a model of statistical analysis as independent variables. For example, the work of Nguyen and colleagues (2016) detect, in three cities, the level of happiness for each census tract through geo-localized tweets. The authors insert the level of happiness as a dependent variable in a model where the independent variables are the socio-economic indicators related to each census tract. In the articles classified as reconstruction, the phenomenon examined is explained through a discursive reconstruction of the socio-economic characteristics of the space. Varicelli (2018), for example, interprets the way in which social classes move within the city starting from the general characteristics of the neighbourhoods and streets. Therefore, between the three categories above, the last two, and in particular the last one, follow the ecological tradition, while the first one represents a simple attempt to describe the geography of the phenomenon analyzed. The second indicator allows us to understand which characteristics of space are considered relevant in the articles classified as "explanatory" or "reconstruction". As we can see in table 4, in half the articles that fall into these two categories (17/34) the variability of the phenomenon analyzed is traced back to the demographic and economic characteristics of the population that resides in the ecological unit under consideration.

Table 4 – Mechanism that relates spatial characteristics and phenomenon analyzed

Mechanism	**n. of article**
Social interaction	17
Environmental/geographic	23
Total	34

This approach to the analysis is in line with the tradition of studies oriented towards the analysis of structural effects (Blau, 1960) and assumes that the analyzed phenomenon is influenced by so-called social-interaction mechanisms (i.e., the product of social interactions between individuals). 6 out of 17 articles, in addition to the characteristics of the population, consider other aspects, such as the characteristics of the housing market; the communicative environment, or aspects related to the distance from the city center and mobility. Considering variables such as real estate, quality of transport and distance from the center, in fact, means to take into account more mechanisms than just the interactive-social one, such as environmental and geographical mechanisms.

Environmental mechanisms refer to the attributes of space that can directly or indirectly influence the behavior of individuals. For example, degraded physical conditions of the buildings can influence both the perceptions and the behavior of residents in different ways.

Geographical mechanisms, on the other hand, refer to aspects of spaces that may influence the life courses of residents, these mechanisms are not endogenous but arise due to the position of the neighborhood in relation to the external environment. Spatial maladjustment, for example, seems to occur when a lack of accessibility to certain services or infrastructures impacts certain social phenomena. Considering several mechanisms simultaneously is in line with the literature on neighborhood effects.

The remaining 17 articles base their analysis on environmental and/or geographical mechanisms, so the characteristics of the ecological unit, such as the type of function performed (i.e. residential; economic, historic center, etc.) or the location of certain types of activities and urban hubs, are considered independent variables. The last indicator concerns the analysis strategy selected, which depends on the unit of analysis considered. In the first strategy, which has also been defined as “areal” (Johnston, 1976), both the independent and dependent variables have the ecological unit as their unit of analysis. In the second strategy of analysis, usually defined as contextual, the independent variables are collected on both the ecological unit and the individual while the dependent variables are collected at the individual level. Although similar, the two analysis strategies show some differences. It is important to highlight that contextual analysis allows inference at the individual level without problems of ecological fallacy. Inference at the individual level is not possible in areal studies, which have other objectives (Menzel,1950).

Lastly, the main problem of contextual analysis is called “misspecification of the model at the individual level”, we find this kind of problem when the observed “contextual or group effect” is due to the omission of individual-level variables related to the outcome and to the group characteristic investigated. In our sample, only 5 studies used the contextual strategy while the remaining used the areal strategy.

5.3 *How to analyze urban geo-data: Tools and issues*

In this section we will discuss the main issues regarding Twitter big data analysis in identifying working algorithms and techniques that might be used. These are useful strategies to answer different questions, though these are only part of all the strategies available. To offer better understanding and reflexive approach to our results some important features of our sample need to be considered. The first is methodological awareness, which regards the description of limitation about extraction and elaboration methods offered by the article authors. This critical approach does not mean authors take into consideration each of the points shown above, nor do they study one of those in depth, but they are interested in highlighting, as far as their research query is concerned, the point that can affect their results. In our sample of articles 35 out of 63 articles pay attention to methodological awareness describing information critically or claiming the position of the authors with respect to the system of data extraction used. A second important point regards the strategies that were used to identify and collect geo-tweets. Geotagging was implemented by 59 of 63 authors, while geoparsing in only 4 articles and none used geocoding. We have chosen the most commonly used tools from the authors of our sample of articles to work with big data pointing out their functioning and how to apply them.

Tab. 5 – Summary: Tools and Issues to make research thanks to Twitter geo-spatial data. Evidences from our sample

TOOLS	ISSUES	REFERENCES
Moran's	What kind of relationship do the objects of research have with each other? Is it due to cause or effects?	Moran 1950
Density	How does density change of particular phenomenon over a period of time?	Sulis *et al* 2018
Spatial Lag	Which are the variables that can explain particular phenomenon?	Chi, Zhu 2008
K-means	How can homogeneous groups be identified while reducing clustering errors?	Likas *et al* 2002
SOM/Geo-SOM	How can clusters be created thanks to a neuronal map in a bi or tridimensional space?	Chaudhary 2014; Bação *et al* 2004
DBSCAN	How can objects be clustered thanks to their density?	Ester *et al* 1996
Girvan–Newman	Which kind of communities are there? How can they be discovered and compared?	Yuan. *et al* 2020

If the aim of the study is to answer the following questions: how do some variables of particular phenomena spread in a determinate geographic area? What are the kinds of relationships between the objects? Is it random, casual or due to effects? Are the variables regarding particular phenomena dispersive or concentrated? Moran's autocorrelation can be used. Moran's index, (Moran, 1950) or rather Moran's I, is defined as a spatial autocorrelation measure to create clusters of similarities while overcoming the assumption of statistical independence, measuring the distance from different objects. It is multidimensional and multidirectional and its range, similarly to correlation, varies from -1 to + 1 including 0 as absence of autocorrelation and the presence of random phenomenon. In addition, it can take different dimension of global or local interest. Other techniques might be used to lead spatial analysis and identify diversity: intensity, variability and consistency. For instance, urban data of mobility can be used to detect the vitality of a city (Sulis *et al* 2018). Intensity helps to identify the concentration of a phenomenon in a defined period; variability, on the other hand, allows us to study differences over a period of time, such as per week; finally, consistency describes the variation in different phases of the same day, such as for per hour. Spatial lag (SL) regression (Chi, Zhu, 2008) is a tool to identify significant explanatory variables particularly regarding crime occurrence (Ristea *et al* 2018).

These techniques were used in our sample to investigate different phenomena - from the traces that people leave during unexpected events (Gulnerman *et al* 2020) to neighborhood characteristics (Anselin, Williams, 2015) or the distribution of some kind of capital (Kiuru, Inkinen, 2019). In addition, in our sample this is sometimes combined with GeoDa software. This is user friendly software that was designed for beginners who want to use spatial analysis with a simple graphic interface.

The second large group is represented by clustering algorithms, they help answer the following questions: how can the concentration of tweets in a determined area allow us to understand the different features, opinion or feelings about one particular phenomenon? Or rather it investigates the differences between and within the clusters themselves regarding some kinds of variables or dimension of the query. K-means is a clustering algorithm that reduces the clustering errors in identifying homogeneous groups (Likas, 2008). K-means has some problems depending on its structure: it identifies clusters starting from the center of the cluster whose position is arbitrarily decided, gradually minimizing the error of the other objects that surround the center. Nevertheless, it is used a lot as a clustering algorithm and some development as for Global K-means were implemented (Likas *et al*., 2002). In our sample it is used sometimes in combination with a self-organizing map (SOM) consisting in a *neural map,* the node of which are called *neurons*. A neural map consists of a 2 or 3- dimensional grid of nodes from multidimensional inputs (Chaudhary, 2014). A further development of SOM that combines SOM and Geographical representation is Geo-SOM (Bação *et al*., 2004). These two algorithms display and map clusters in a geographical way, providing a graphic representation of data. Another clustering system is Density-Based Spatial Clustering of Applications with Noise (DBSCAN), which is a density-based algorithm (Ester *et al*., 1996). In this algorithm two points can be clustered if they are less distant than a predetermined distance, if the cluster is not created, it will be defined as rumors. Another way to identify *communities* -or clusters- is the Girvan–Newman algorithm. This is a hierarchical method of detecting communities in complex networks, which can also be applied to weighted networks. Within each step of the Girvan–Newman community detection algorithm, it progressively removes edges with highest edge betweenness centrality (i.e., edges with highest number of shortest paths passing through them) and recalculates edge betweenness after each iteration of removal. By removing these high betweenness edges, the communities are separated from each other and consequently, the underlying community structure of the network is revealed (Yuan., Crooks., *et al*., 2020). In this way, disconnected components are created within the network that identifies the communities present in it. As for unweighted networks, the algorithm removes the edges of maximum interference whose final result will be the dendrogram that repeats the steps until no more edges can be removed or more ideal communities have been reached (i.e., the highest modularity of the clusters). For example, in the work by Yuan X., Crooks A. e Züfle A. (2020), the authors used the Girvan–Newman algorithm to conduct network community detection (i.e., clustering) on the sparsified networks. So, after thematic similarity networks are constructed, the authors carried out network community detection. A common evaluation metric for network community detection quality is modularity, which is a measure of the strength of division of a network into communities (Yuan., Crooks., *et al*., 2020). In our sample the tools described above were used, for instance: to identify the power of social media data in modern cities (Arribas-Bel *et al*., 2015) or traffic dynamics when disruption occurred (Steiger *et al* 2016) or even tourist movement patterns (Hu *et al*., 2018).

Conclusion

The work presented in this paper concerns the analysis of the spatial dimension, considered to be particularly promising in the field of urban analysis (Singleton *et al*., 2018) and in the emerging trend of digital studies following the effective spread of the use of social media in cities.

Space, in addition to being conceived in a physical sense and in relation to a specific phenomenon, has been considered in relation to the digital world and in particular with social media, specifically considering geographical information from Twitter.

In order to investigate the main features of this line of emerging studies a systematic review of the literature based on the analysis of the content of a sample of articles extracted from the Scopus database was carried out with two main objectives: 1) to analyze the central points such as the topics under investigation; emerging concepts of space; awareness of the limits of geolocalized data and of the spatial statistics techniques used, and 2) to provide some coordinates to be followed to set up the searches with the geographic data.

The systematic review was carried out through the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-analysis) methodology, which shows all the selection steps during the analysis of the articles.

The themes that emerged from the full papers considered were divided into six clusters. This clusters contain papers that apply the geospatial Twitter data to the most diversified social phenomena such as social movements, mobility, the use of urban spaces or transport etc.

For further developments of this research, it would be interesting to consider the works that use geospatial data in relation to the Covid 19 pandemic from this analysis, another result regards the conception of space. What emerged in most of the articles analyzed is the identification of a sociological conception of space, while a residual number of articles seem to be characterized by a data driven perspective. In fact, in many articles classified as "descriptive" sociological space is replaced by other valid factors that are useful for answering the research questions. In the articles where a sociological conception of space prevails, different kinds of mechanisms are taken into account to explain the phenomenon analyzed. In the next phases of the research, it will be interesting to analyze which of these mechanisms have a significant impact on the communication phenomena detected through twitter and in the same way to understand how the authors open the "black box" underlying the statistical evidence. This evidence sometimes leads to undue generalizations, in fact one of the problems related to the ecological approach is the so-called ecological fallacy. It will be interesting to see if this error is still present in the works that take space into consideration. Finally, we chose to consider some research questions that spatial analysis tools are able to answer. Our sample of articles can be used as reliable guidelines to answer identified research questions. As far as the use of tools is concerned, one important issue regards the relationships between issues and tools. Some kinds of topic are related to a particular tool of analysis. Furthermore, particular attention to the whole research process is needed. In particular, this refers to the strategy of extraction and use of geo-localized tweets that have to include methodological awareness, in order to avoid "data speaks from itself" that could lead to simplistic analysis of results. According to the aims of this paper, the authors suggest that: the use of geo-data from Twitter must include a careful analysis of each passage from the query to the interpretation of results, taking into account the limitation described above. The results of the analysis show, on one hand, the complexity of this important corpus of research, and on the other, this can be an introductory tool for the neophyte. Which are the topics to investigate? What kind of geo-referenced data exist? What are the advantages and disadvantages in using these? What are the possible perspectives of the relationship between data and territory? Which tools can I use to answer to my issues? These are the questions that we have tried to answer, and which will be useful in this kind of analysis. As we claimed above, our research was based on Scopus, the best platform for social science. However, other platforms can be used taking into consideration the different features of subject sub-categorization. Further research can explore the relationship between methodological and theoretical approaches, improving the level of abstraction of our categories to other kinds of field analysis of space.

References

Anselin, L., Williams, S. (2015). Digital neighborhoods. *Journal of Urbanism: International Research on Placemaking and Urban Sustainability*, 9(4), 305–328. doi:10.1080/17549175.2015.1080752.
Archer, J., Pattie, C. (1998). Cox, K.R. 1969: The voting decision in a spatial context. Progress in Geography, 1, 81-117, *Progress in Human Geography*, 22, 407-410.
Arribas-Bel, D., Kourtit, K., Nijkamp, P. *et al*. (2015). Cyber Cities: Social Media as a Tool for Understanding Cities, *Applied Spatial Analysis and Policy*, 8 (3). 231 - 247., doi.org/10.1007/s12061-015-9154-2.
Bação F. Lobo V., Marco Painho M. (2004). Geo-Self-Organizing Map (Geo-SOM) for Building and Exploring Homogeneous Region, Conference: *Geographic Information Science, Third International Conference, GIScience* 2004, Adelphi, MD, USA, October 20-23, Proceedings.
Bagnasco A. Les Galès P. (2000). European cities: local societies and collective actors? In Arnaldo Bagnasco, Patrick Les Galès, *Cities in Contemporary Europe,* Cambridge: Cambridge University Press.
Bakerman J., Pazdernik K., Wilson A., Fairchild G., Bahran R (2018). Twitter Geolocation: A Hybrid Approach, *ACM Transactions on Knowledge Discovery from Data*, Vol. 12, No. 3, Article 34, doi.org/10.1145/3178112.
Berelson, B. (1952). *Content Analysis in Communication Research*. Michigan: Free Press.
Blalock, H. M. (1984). Contextual-effects models: theoretical and methodological issues, *Annual review of sociology*, *10*(1), 353-372, doi.org/10.1146/annurev.so.10.080184.002033.
Blau, P. M. (1960). Structural effects, *American sociological review*, 25, 178–193. doi.org/10.2307/2092624.
Burrell, J. (2016), How the machine 'thinks': Understanding opacity in machine learning algorithms*, Big Data Society*, vol.1, n.12, doi.org/10.1177/2053951715622512.
Butler, D., Stoke, D., (1974). *Political change in Britain: Basis of electoral choice,* Berlino: Springer.
Campagna M., (2014). The Geographic Turn in Social Media: Opportunities for Spatial Planning and Geodesign, in Murgante B. *et al*. (ed*.), Computational science and its applications* – ICCSA, pp. 598-610, Berlino: Springer, DOI: 10.1007/978-3-319-09129-7_43.
Chaudhary V., Bhatia R. S., Ahlawat A. K. (2014). A novel Self-Organizing Map (SOM) learning algorithm with nearest and farthest neurons, *Alexandria Engineering Journal*, doi.org/10.1016/j.aej.2014.09.007.
Chi, G., Zhu, J. (2008). Spatial Regression Models for Demographic Analysis, *Population Research and Policy Review*, 27(1), 17-42, DOI 10.1007/s11113-007-9051-8.
Courmont, A. (2021). *Quand la donnée arrive en ville: open data et gouvernance urbaine. Open data et gouvernance urbaine,* Grenoble : Pug Presses Univer De Grenoble.
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*, 24 (4), DOI: 10.1080/13658816.2014.996567.
De Vos, S. (1998). The analysis of compositional effects as exemplified by the study of elections. *GeoJournal*, *44*(1), 43-49, doi.org/10.1023/A:1006820202876.
Dongo I. Cadinale Y. Aguilera A. (2020). Web Scraping versus Twitter API: A Comparison for a Credibility Analysis, *iWAS '20: The 22nd International Conference on Information Integration and Web-based Applications Services*, Chiang Mai, Thailand, Association for Computing Machinery: New York doi.org/10.1145/3428757.3429104.
Dourish, P. (2016), Algorithms and their others: Algorithmic culture in context, *Big Data & Society*, vol 1, n.11, doi.org/10.1177/2053951716665128.
Durkheim, E., (1893 [1962]). *La divisione del lavoro sociale*, Milano: Edizioni di comunità.
E. Menduni (2014), *I media digitali: tecnologie, linguaggi, usi sociali*, Bari: Laterza.
Economist. (2012) The New Local, October 27, p. 14, www.economist.com/special-report/2012/10/27/the-new-local.
Edwards, A., *et al*l., Digital social research, social media and the sociological imagination: surrogacy, augmentation and re-orientation in *International Journal of Social Research Methodology*, V. 16, 2013, Issue 3, doi.org/10.1080/13645579.2013.774185
Ester M., Kriegel H. P., Sander J. Xu X. (1996). A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise, *KDD'96: Proceedings of the Second International Conference on Knowledge Discovery and Data Mining,* 226–231.
Fei Hu, Zhenlong Li, Chaowei Yang, Yongyao Jiang (2018): A graph-based approach to detecting tourist movement patterns using social media data, *Cartography and Geographic Information Science*, 46(3) DOI: 10.1080/15230406.2018.1496036.
Fu C., McKenzie G., *et al*l. (2018). Identifying spatiotemporal urban activities through linguistic signatures, *Computers, Environment and Urban Systems*, doi.org/10.1016/j.compenvurbsys.2018.07.003.
Galster, G. C., (2012). The mechanism (s) of neighbourhood effects: Theory, evidence, and policy implications, In Van Ham, M., Manley, D., Bailey, N., Simpson, L., Maclennan, D., *Neighbourhood effects research: New perspectives* (pp. 23-56), Dordrecht: Springer.
Garcia A.C *et al*. (2009), Ethongraphic Approaches to the Inernet and Computer-Mediated Community,*Journal of Contemporary Ethnography*, 38, pp. 52-84, doi.org/10.1177/0891241607310839.
Gardini, E. (2010). La dimensione dello spazio. Per una sociologia spazialista, *Sociologia urbana e rurale*, V.91, FrancoAngeli: Milano, doi10.3280/SUR2010-091006.

Generous, N., Fairchild, G., Deshpande, A., Del Valle, S. Y., Priedhorsky, R. (2014). Global disease monitoring and forecasting with Wikipedia, *PLoS Comput Biol*, *10*(11), doi.org/10.1371/journal.pcbi.1003892.
Giddens, A. (1979). *Central problems in social theory: Action, structure, and contradiction in social analysis*, London: The Macmillan Press.
Giddens, A., (1981). *A Contemporary Critique of Historical Materialism*, London: MacMillan.
Gieryn, T., F., (2000). A Space for Place, *Sociology Annual Review of Sociology*, 26, 463–96, doi.org/10.1146/annurev.soc.26.1.463.
Gobo, G., Mauceri, S. (2014). *Constructing survey data: An interactional approach*, New York: Sage.
Goodchild, M. F., (2007) *Citizens as sensors: the world of volunteered geography*, GeoJournal, 69(4), 211-221. DOI 10.1007/s10708-007-9111-y.
Gritta, M., Pilehvar, M.T. Collier, N. A pragmatic guide to geoparsing evaluation. *Lang Resources & Evaluation*, 54, 683–712 (2020), doi.org/10.1007/s10579-019-09475-3.
Guhathakurta S., Zhang G., Chen G., Burnette C., Sepkowitz I.(2019). Mining Social Media to Measure Neighborhood Quality in the City of Atlanta, *International Journal of E-Planning Research*, 8(1) DOI: 10.4018/IJEPR.2019010101.
Gulnerman A.G., Karaman H., Pekaslan D., Bilgi S. (2020). Citizens' Spatial Footprint on Twitter—Anomaly, Trend and Bias Investigation in Istanbul, *ISPRS Int. J. Geo-Inf.* 9(4), 222, doi.org/10.3390/ijgi9040222.
Menzel H., (1950). Comment on Robinson's Ecological Correlation and Behavior of Individuals, *American Sociological Review*, 15, p. 674, doi.org/10.2307/2087176.
Haffner, M., (2018). A place-based analysis of #BlackLivesMatter and counter-protest content on Twitter, *GeoJournal*, V.84, 2020, doi:10.1007/s10708-018-9919-7
Harvey, D. (1978*). Giustizia sociale e città* (2 voll.). Milano: Feltrinelli.
Harvey, D. (1993). From Space to Place and Back Again: Reflections on the Condition of Postmodernity. In J. *et al*. Bird (Ed.), *Mapping the Futures: Local Cultures, Global Change*, London: Routledge.
Helmes-Hayes, R. (1987). "A Dualistic Vision": Robert Ezra Park and the Classical Ecological Theory of Social Inequality. *The Sociological Quarterly, 28*(3), 387-409, doi.org/10.1111/j.1533-8525.1987.tb00302.x
Hu F., Li Z., Yang C., Jiang Y. (2018). A graph-based approach to detecting tourist movement patterns using social media data, *Cartography and Geographic Information Science*, DOI: 10.1080/15230406.2018.1496036.
Jacobs, J. (1961). *The death and life of great American cities*, New York: Random House.
Jiang, Y., Li, Z., Ye, X. (2019), Understanding demographic and socioeconomic biases of geotagged Twitter users at the county level, Cartography and geographic information science, vol. 46, n.3, 228-242.
Johnston, R. (1976). Areal Studies, Ecological Studies, and Social Patterns in Cities, *Transactions of the Institute of British Geographers*, 1(1), 118-122. doi:10.2307/621317
Kazepov Y. (2005). *Cities of Europe. Changing contexts, local arrangements, and the challenge to urban cohesion*, Hoboken: Blackwell Publishing.
Kitchin R. (2014). The data revolution: Big data, open data, data infrastructures and their consequences, *Journal of Regional Science*, 56(4): 722-723, DOI: 10.1111/jors.12293.
Kiuru J., Inkinen T. (2019). E-Capital and Economic Growth in European Metropolitan Areas: Applying Social Media Messaging in Technology-Based Urban Analysis, *Journal of Urban Technology*, 26:2, 67-88, DOI: 10.1080/10630732.2019.1579513
Korkmaz, G., Cadena, J., Kuhlman, C. J., Marathe, A., Vullikanti, A., Ramakrishnan, N. (2015*)*. Combining heterogeneous data sources for civil unrest forecasting, *Proceedings of the 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining* (pp. 258-265).
Kovacs-Györi, A., Ristea, A., Kolcsar, R., Resch, B., Crivellari, A., Blaschke, T. (2018). Beyond Spatial Proximity—Classifying Parks and Their Visitors in London Based on Spatiotemporal and Sentiment Analysis of Twitter Data, *ISPRS International Journal of Geo-Information*, 7(9), 378. doi:10.3390/ijgi7090378
Krippendorff, K. (2018). *Content analysis: An introduction to its methodology*. New York: Sage publications.
Lansley G., Longley P.A. (2016). The geography of Twitter topics in London, *Computers, Environment and Urban Systems*, 58:85-96, doi.org/10.1016/j.compenvurbsys.2016.04.002.
Lazarsfeld, P. F., Menzel, H. (1961). On the relation between individual and collective properties. *Complex organizations: A sociological reader*, 422-440.
Li, Y., Li, Q., Shan, J. (2017). Discover Patterns and Mobility of Twitter Users—A Study of Four US College Cities, *ISPRS International Journal of Geo-Information*, 6(2), 42. doi:10.3390/ijgi6020042.
Li, Z., Wang, C., Emrich, C. T., Guo, D. (2017). A novel approach to leveraging social media for rapid flood mapping: a case study of the 2015 South Carolina floods*, Cartography and Geographic Information Science, 4*5(2), 97–110*,* doi:10.1080/15230406.2016.1271356.
Likas A., (2008). The Global Kernel k-Means Clustering Algorithm, *EE Transactions on Neural Networks 20(7):1181-94*, DOI: 10.1109/IJCNN.2008.4634069.
Lutters, W. G., Ackerman, M. S. (1996). An introduction to the Chicago School of Sociology. Interval Research Proprietary, 2(6), 1-25.
Malik, M., Lamba, H., Nakos, C., Pfeffer, J. (2015). Population Bias in Geotagged Tweets, *Proceedings of the International AAAI Conference on Web and Social Media*, 9(1).

Massa, P., Campagna M., 2014, Social Media Geographic Information: Current Developments and Opportunities in Urban and Regional Planning, *REAL CORP 2014 – PLAN IT SMART! Clever Solutions for Smart Cities. Proceedings of 19th International Conference on Urban Planning, Regional Development and Information Society,* 631-640
Mela, A. (2006). *Sociologia delle città*, Roma: Carocci.
Meyer T.R., Balagué D., Camacho-Collados M., Li H., Khuu K., Brantingham P.J., Bertozzi A.L.(2019). A year in Madrid as described through the analysis of geotagged Twitter data, *Environment and Planning B: Urban Analytics and City Science*, 46, 9, doi.org/10.1177/2399808318764123.
Middleton, S. E., Kordopatis-Zilos, G., Papadopoulos, S., Kompatsiaris, Y. (2018), *Location extraction from social media: Geoparsing, location disambiguation, and geotagging*, ACM Transactions on Information Systems, TOIS, 36(4), 1-27.
Moran P.A.P. (1950). Notes on Continuous Stochastic Phenomena, *Biometrika*, Jun., 1950, Vol. 37, No. 1/, pp. 17-23, doi.org/10.2307/2332142.
Natale, P., Airoldi, M., (2017). *Web & Social Media. Le tecniche di analisi*, Santarcangelo di Romagn: Maggioli Editore.
Nguyen, Q. C., Kath, S., Meng, H. W., Li, D., Smith, K. R., VanDerslice, J. A., Wen, M. Li, (2016), Leveraging geotagged Twitter data to examine neighborhood happiness, diet, and physical activity*, Applied Geography*, vol. 73, 77-88, DOI: 10.1016/j.apgeog.2016.06.003.
Noulas A., Scellato S., Lathia N. and Mascolo C., (2012), Mining User Mobility Features for Next Place Prediction in Location-Based Services, *IEEE 12th International Conference on Data Mining, Brussels*, 2012, pp. 1038-1043.
Park, R., Burgess, E., W., (1970), *Introduction to the Science of Sociology: Including the Original Index to basic Sociological Concepts,* Chicago: University of Chicago Press.
Pintaldi, F. (2000). Alla ricerca dell'inferenza perduta: la relazione tra dati ecologici e dati individuali*, Sociologia e ricerca sociale,* 63/2000, Doi
Rabari C., Storper M., (2015). The digital skin of cities: urban theory and research in the age of the sensored and metered city, ubiquitous computing and big data*, Cambridge Journal of Regions, Economy and Society*, 8, 27–42.
Ristea A., Andresen M.A., Leitner M., (2018). Using tweets to understand changes in the spatial crime distribution for hockey events in Vancouver, *The Canadian Geographer*, 62(3), pp. 338–351, DOI: 10.1111/cag.12463.
Robinson (1950). Ecological correlations and the behavior of individuals, *American Sociological Review*, 15(3), 351-357, doi.org/10.2307/2087176.
Robinson, P., Lowe, J. (2015). Literature reviews vs systematic reviews, *Australian and New Zealand journal of public health*, *39*(2), 103-103, doi: 10.1111/1753-6405.12393.
Rogers, Richard., (2009). *The end of the Virtual: Digital Methods*, Amsterdam: Amsterdam University press.
Sassen, S., (1994) *Cities in a World Economy*, London: Sage.
Shi, Y., Deng, M., Yang, X., Liu, Q., Zhao, L., Lu, C.-T. (2016). A Framework for Discovering Evolving Domain Related Spatio-Temporal Patterns in Twitter*, ISPRS International Journal of Geo-Information*, 5(10), 19, doi.org/10.3390/ijgi5100193.
Simmel G., (1903 [1999]). Le metropoli e la vita dello spirito, Roma : Armando Editore
Sizov S. (2012). *Latent geospatial semantics of social media,* ACM Transactions on Intelligent Systems and Technology (TIST), 3, 1-64:20.
Stefanidis A., Crooks A., Radzikowski J., (2013). Harvesting ambient geospatial information from social media feeds, *GeoJournal,* Vol.78, pp. 319–338, doi.org/10.1007/s10708-011-9438-2
Steiger E., Resch B., de Albuquerque J.P., Zipf A. (2016). Mining and correlating traffic events from human sensor observations with official transport data using self-organizing maps, *Transportation Research Part C: Emerging Technologies*, 73, 91-104, doi.org/10.1016/j.trc.2016.10.010.
Strassoldo, R. (1990). The Sociology of Space, *Culture-Space-History: Proceedings 11th International Conference of the IAPS,* Ankara, Turkey.
Sulis P., Manley E., Zhong C., Batty M., (2018). Using mobility data as proxy for measuring urban vitality, *Journal of spatial information science*, 16 (2018), pp. 137–162, DOI: 10.5311/JOSIS.2018.16.384.
Taylor L., Richter C., (2015). Big Data and Urban Governance, in Gupta J., Pfeffer K., Verrest H., Ros-Tonen M. (2015). *Geographies of Urban Governance Advanced Theories, Methods and Practices,* Switzerland: Springer.
Tranfield, D., Denyer, D., Smart, P. (2003). Towards a methodology for developing evidence-informed management knowledge by means of systematic review, British *journal of management*, 14(3), 207-222, DOI: 10.1111/1467-8551.00375.
Van Dijck J, De Waal M, Poell T. (2018). The Platform Society: Public Values in a Connective World, Oxford, Oxford University Press Inc
Van Ham, M., Manley, D., Bailey, N., Simpson, L., Maclennan, D. (2012). Neighbourhood effects research: New perspectives. In *Neighbourhood effects research: New perspectives* (pp. 1-21). Springer, Dordrecht.
Vicari Haddock S., (2013). *Questioni urbane. Caratteri e problemi della città contemporanea*, Bologna: il Mulino.
Weber M. (1914 [2003]) *Economia e società. La città*, Roma: Donzelli.
Weber, R. P. (1990). *Basic content analysis* (No. 49). New York: Sage.
Wilson, W. J. (1987). *The truly disadvantaged: The inner city, the underclass, and public policy,* University of Chicago Press: Chicago.

Wu F., Li Z., Lee W.-C., Wang, H., Huang, Z. (2015), Semantic annotation of mobility data using social media, *Proceedings of the 24th international conference on World Wide Web. International World Wide Web Conferences Steering Committee*, Republic and Canton of Geneva 1253–1263, doi.org/10.1145/2736277.2741675.

Yuan, X., Crooks, A., Züfle, A. (2020). A Thematic Similarity Network Approach for Analysis of Places Using Volunteered Geographic Information, *ISPRS International Journal of Geo-Information*, 9(6), 385. doi:10.3390/ijgi9060385

Zhang W., Gelernter J. (2014). Geocoding location expressions in Twitter messages: A preference learning method, *Journal of spatial information science*, Number 9 (2014), pp. 37–70, DOI: 10.5311/JOSIS.2014.9.170.

Zhang, G., Zhu, A. X. (2018). The representativeness and spatial bias of volunteered geographic information: a review, *Annals of GIS*, V. 24, n.3, 151-162, doi.org/10.1080/19475683.2018.1501607.

Zhou X, Xu C. (2017). Tracing the Spatial-temporal evolution of Events based on Social media data, *ISPRS Int. J. Geo-Inf.*, 6(3), 88; doi.org/10.3390/ijgi6030088.

Zhou, X., Chen, L. (2014). Event detection over twitter social media streams. *The VLDB journal*, *23*(3), 381-400, DOI: 10.1007/s00778-013-0320-3.

Zupi, M., (2017). Social GIS per l'analisi dei comportamenti e delle abitudini in ambito urbano, *Discovering the world through Gis, VIII Edizione, Catanzaro*.