Follow the Geographic Information:
The Challenges of Spatial Analysis in Digital Methods

*By Ciro Clemente De Falco[*], Gabriella Punziano[±] &
Domenico Trezza[°]*

In the digital environment, defined as a space with no anchors (Menduni 2014), the spatial dimension may have a significant role, mostly in relation to the Internet or digital studies. Social Media Geographic Information (Campagna et al. 2016), even if limited, can be highly useful to overcome some limitations of social media analysis and user generated content. With the objective to report the potential and limits of this approach, in this article a case study will be presented on the individual perception related to COVID-19 in Italy. Starting from the analysis of ecological data and the social media data, we reproduce three measures by exploring three different elements that could be anchored to a geographical dimension: the storytelling of COVID (COVID-Issues), the spread of infection (COVID-Spread), and the distribution of measures (COVID-Measures) to understand the regional trends of the second wave of the pandemic emergence that represent the timespan of this contribution.

Introduction:
On Space, Sociology, and the Study of Digital Diffusive Phenomena

The Role of Space Category in Sociological Studies

Space and its characteristics in the history of social theory and in the sociology of the twentieth century have not always played a central role. For example, Parson in his first elaborations argued that space was an irrelevant aspect for the analysis of social action. Due to this, some authors have critically spoken of sociology as an a-spatial discipline (Mela 2006). However, space has played a leading role in many works of classical authors such as Durkheim (Halbwachs 2018), Simmel (Cotesta et al. 2010) and Park (Mannella 2009). Durkheim, indeed, included space in his (never started) research program on "social morphology"; Simmel stated that space represented an integral part of his study of "social forms" (Frisby 2002), while in Park's work "human ecology", although with peaks biological determinism, space was one of the main dimensions. Nevertheless, in the recent debate, some relevant authors have tried to put the spatial dimension back into the core of social sciences. For example, in his project for the re-foundation of social knowledge, Giddens (1987) assigns a central role to the need

---

[*]Research Fellow, University of Naples Federico II, Italy.
[±]Assistant Professor, University of Naples Federico II, Italy.
[°]Research Fellow, University of Naples Federico II, Italy.

to reconceptualize the role played by the notions of time and space into social change, and to overcome traditional disciplinary boundaries.

Although with different views and shades, the underlying idea of the cited authors is the belief that space, understood as a sort of habitat, is the result of a process of social production. Furthermore, space should be considered a dimension that can generate and even influence social phenomena.

Recently, thanks to big data seem to be renewed interest in the spatial dimension. In the data revolution era (Kitchin 2014), in fact, new data and new sources allow researchers to find new ways to study society and its dynamics. In particular, geo-located data enable better ways of producing social knowledge (Halford et al. 2013).

Thanks to the spread of smartphones and mobile devices, a class of geo-social software applications that integrates location has emerged. Twitter, Facebook, Instagram, Foursquare, as well as other social platforms, enable, through users' actions, the production of spatial datasets in which we can detect and locate individual perceptions, interactions, and experiences into physical space. This wide availability of geo-localized data led to the development of an approach, whose aim is to jointly analyze two worlds previously considered without meeting points: the online and the offline dimension. It becomes clear, in fact, that the virtual and material separation between the above dimensions, as is highlighted in spatial mediation theory (Leszczynski 2015), is entirely artificial.

This new space conception has emerged thanks to two new kinds of data: *Volunteered Geographic Information* (Goodchild 2007) and *Geographic Information deriving from Social Media* (Campagna et al. 2016). This kind of data shows higher spatial and temporal resolutions than the conventional data sources (i.e., census data and surveys; Batty 2013). Thus, the spatial dimension is becoming the object of increasing attention also in the context of the Internet and digital studies thanks to the availability of these geo-localized data. Chappell et al. (2017) argue that thanks to this kind of data, innovative methods can be developed to study social phenomena and to help sociology to avoid the "incoming crisis" (Burrows and Savage 2014) resulting from the increase of "social" data users.

Five Waves in Social Research on Geo-Social Media Data

The spatial dimension can indeed be declined in multiple ways as well as being conceived in a physical or non-physical way. There are at least five lines of research in the growing research stream that uses geo-social media.

In the first research line, there are works, that use geolocated data coming from social media, to explain how better manage different kinds of situations (i.e., emergency management). Event detection algorithms (Nurwidyantoro and Winarko 2013) integrated into social media monitoring systems use geo-located data to identify places hit by catastrophic events in order to intervene quickly. In addition, event detection algorithms are useful to identify emerging dynamics within the city that require immediate action (Wei et al. 2018).

The second research line aims to analyze the geographical characteristics of some social phenomena, such as the distribution of ethnic groups in large cities

(Longley et al. 2015) or linguistic diversity (Zhao and Cao 2017). Both these research streams are mainly descriptive and not always show a systematic integration of the study of the online world with the offline one.

This integration becomes central in the third and fourth research lines, where there are even more sophisticated theoretical frames. In the third research line, for example, there are mainly qualitative studies showing a focus on understanding how the online world influences the offline world. The more general aim is to understand how digital representations in social media could alter the meaning and the perception of physical environments through visualization and naming, and therefore how the spaces of representation can change spatial practices (Rzeszewski 2018). Sutko and de Souza Silva (2011, p. 810) who investigate the connections between the social and the spatial through geosocial applications and services and their impact on the social production of space and the spatial production of society, underline the transformation of some relational dynamics, such as sociability analyzed and described by Simmel.

In the fourth research line, there are mainly quantitative works, and the focus is explaining the variability of the phenomena investigated through statistical models in which socio-economic variables are considered independent variables. Thus, this growing body of research which works at multiple levels of geographical detail, investigates the different ways in which the spatial dimension is related to what happens in the online world and in particular on social platforms. The last two research streams carry on the tradition of studies begun with authors such as Durkheim (1951), using the ecological approach to connect and explain social phenomena through their spatialization and territorialization (Zajczyk 1991).

Finally, these four research lines are followed by the fifth concerning the development of techniques to study and analyze the content of geo-located data coming from social media (De Falco et al. 2021).

Operationalize the Category of Space in Geo-Social Media Data Analysis: An Empirical Proposal

In our opinion, one of the most promising for sociological research is the fourth wave exposed to which we linked the main objective of this contribution that consists of reporting the potential and limits of this approach. To this end, this paper will be divided into sections. In the first, through a case study on the individual perception related to COVID-19 in Italy, we will see the potential deriving from the use of the spatial dimension in the analysis of tweets. The analysis, conducted at the regional level, will investigate, thanks to the use of topic modeling the regulatory, contextual, and geographic influences of the spread of the virus on the topics that emerged from them. In the second section, however, we will see what the limits of this analysis are and what dilemmas they pose to the researcher. Since this is an exploratory analysis, no conclusions are drawn in this study. But, we intend to leave the reader with an analytical procedure that will highlight the potential of geo-referenced social data to study particular phenomena and their normative and diffusive components.

Making a Geography of the Storytelling: The Case of COVID-19 and its Widespread Perception in Italy

Once learned the theoretical context in which the relationship between space and digital environment could find interesting developments, it would be useful to understand the empirical implications. Our example of research related to an application on geolocalized tweets on COVID-19 content coming from Italy (well known as one of the European countries most affected by the recent pandemic) in a period covering almost three months, from October the 24th to December 18th, 2020, that is the period corresponding to the second phase/wave of the emergency.

The aim of the research was to see if differences in the geography of infection had implications for the geography of social storytelling as well. The research assumed that the pandemic, at least in its initial phase, did not affect Italy uniformly. In fact, the first epidemic wave had a strong geographical pattern, identified in a very limited area of Lombardy, and then spreading to other regions mainly in Northern Italy. The idea was that due to the shift in the diffusion trend (concentrated in Northern Italy at the beginning and then spreading to the whole country in the second wave), a more cohesive storytelling about negative sentiments would also emerge from the digital arena. This question was addressed by analyzing the corpus of geolocated tweets produced in the second phase of the emergency. The scrape of tweets occurred during the period that goes from the renewed October closures to the partial pre-Christmas re-openings that characterized Italian governance measure of containment of the virus spread during that period.

However, we diversified three normative periods during this time span. The first, corresponding to "general limitations" (24th Oct. – 2nd Nov.), relates to the general measures in the DPCM of October 13th, 2020, reintroducing several general restrictions and the requirement for anti-virus protection. The second normative phase, "traffic lights" (3rd Nov. – 4th Dec.), is related to the introduction of risk scenarios (DPCM of November 3, 2020). For the first time since the pandemic started, it sets out a differentiation of measures for regions, depending on the severity of the epidemiological situation. The third, "Christmas norms" (4th Dec. – 18th Dec.), maintains the earlier instructions on risk ranges and introduces some limitations for the upcoming Christmas holidays.

Tweets were extracted using automated techniques via Twitter's API (Application Programming Interface), initiated in the R environment by "rtweet" package. It allows us to interface with Twitter and set up procedures that fit our extraction standards, including spatial data characteristics. Geolocated tweets were attributed to the region from which they are tweeted. The dataset consists of 11736 tweets, and it has been explored by combining text mining techniques and GIS analysis. This allowed us to map the most recurring themes in social discourse on Twitter. In this way, the thematic element can give rise to a territorial spread corresponding to *COVID-Issues*, the regulatory element can give rise to a territorial differentiation of governmental measures corresponding to *COVID-Measures*, while the diffusive element of virus contagion over the national territory gives rise to a final geographic measure of *COVID-Spread* of the virus.

These are, precisely, three diffusive geographies that can be adequately produced and represented with GIS techniques.

The three emerging geographies, on the storytelling of COVID (*COVID-Issues*), on the spread of infection (*COVID-Spread*), and on the distribution of measures (*COVID-Measures*), were related to understanding the trend of this second wave of the emergence.

First of All: How much is Posted in Each Region?

The potential of geographic information was summarized in a geographical representation of regions with the most tweets. The tweet spatial information allowed us to identify areas with more tweets. The maps allow a visual comparison within the three periods of observation (Figure 1). Depending on the periods, it can be noted how there are regions that are "darker" than others and that are often connected to more relevant crisis conditions. For example, Veneto in the first period is connected to a progressive emergency situation. In the second period, Basilicata, Calabria and Valle d'Aosta (likely due to the controversies connected to the attribution of the color for these areas). For the third period, most of the regions of the Center: Abruzzo, Umbria and Molise (e.g., the controversies related to the situation in Abruzzo, which sees the region in the red zone for a relatively prolonged period).

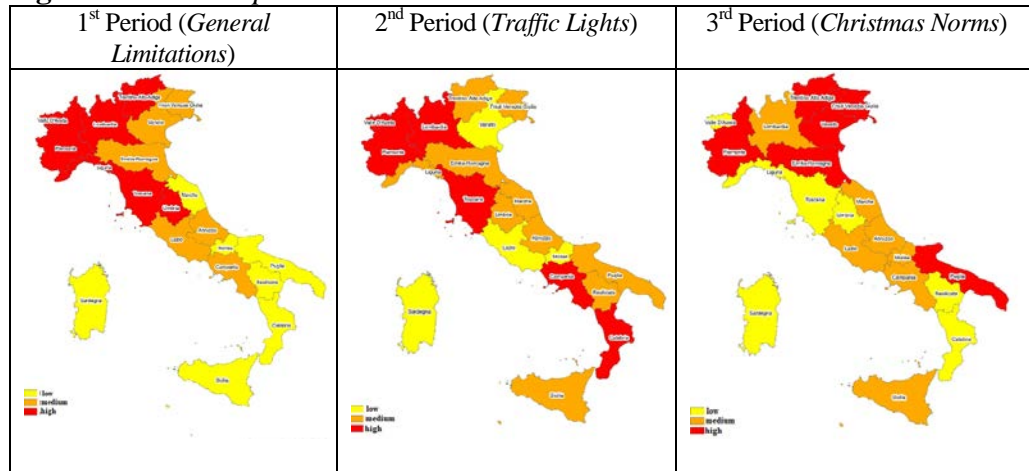**Figure 1.** *Map of Tweets in the Three Periods, % on Tot. per Region*

1ˢᵗ Period (*General Limitations*)	2ⁿᵈ Period (*Traffic Lights*)	3ʳᵈ Period (*Christmas Norms*)
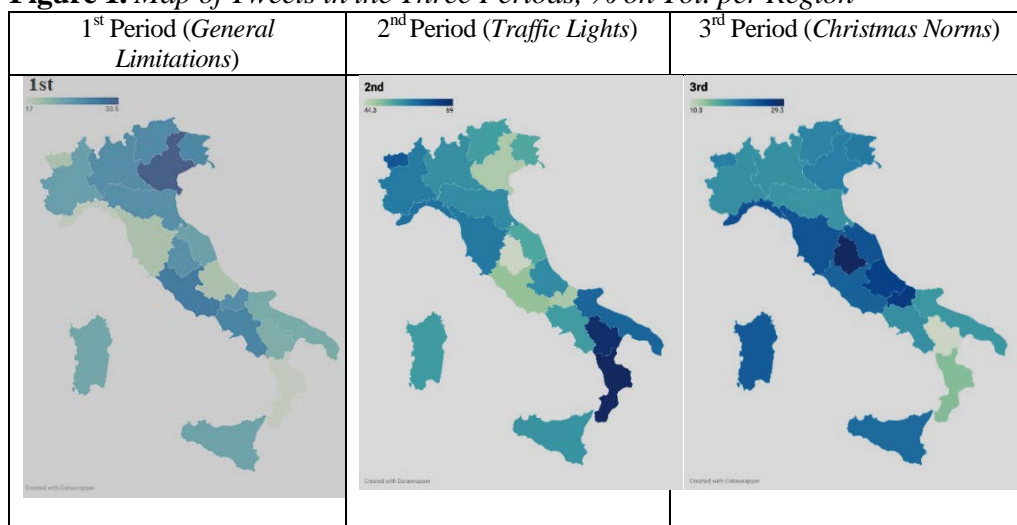
The COVID-Spread Geography

The creation of *COVID-Spread* maps, i.e., the territorial spread of the virus in Italy, required the creation of a variable on the impact of contagion for each region (Figure 2). This is the ratio between the number of infections on the resident population, then the value was perceptualized and categorized on three levels (low, up to 33%, medium, up to 66%, high, for the remaining values up to 100%).

The cartographic representations below provide a summary for each period. It emerges that only in three regions has the impact of disease not changed - Sardinia, Abruzzo and Piedmont - where it has remained low, medium and high respectively. Others, on the other hand, have registered significant variability, in some cases tending to increase (Veneto, Molise, Puglia...), in others tending to decrease (Valle d'Aosta, Umbria, Liguria.). In any case, the overall situation changes a lot, but does not seem to improve significantly. Moreover, the concentration of the Virus remains high in the North of the country for the entire period.

Figure 2. *COVID-Spread*

| 1st Period (*General Limitations*)	2nd Period (*Traffic Lights*)	3rd Period (*Christmas Norms*)

The COVID-Measures Geography

COVID-Measures were categorized on the normative scenario prevalent within the three periods considered (Figure 3). As can be seen from the maps below, there is a trend gap with the impact of the virus, that is, with the likely "real" scenario of the emergency. As opposed to the maps on the *COVID-Spread*, for many regions there is a downward trend, as we go from very stringent regulatory provisions (red zones) for most regions, to a scenario of moderate risk (yellow zones) in the last period considered.

Figure 3. *COVID-Measures*

1st Period (*General Limitations*)	2nd Period (*Traffic Lights*)	3rd Period (*Christmas Norms*)
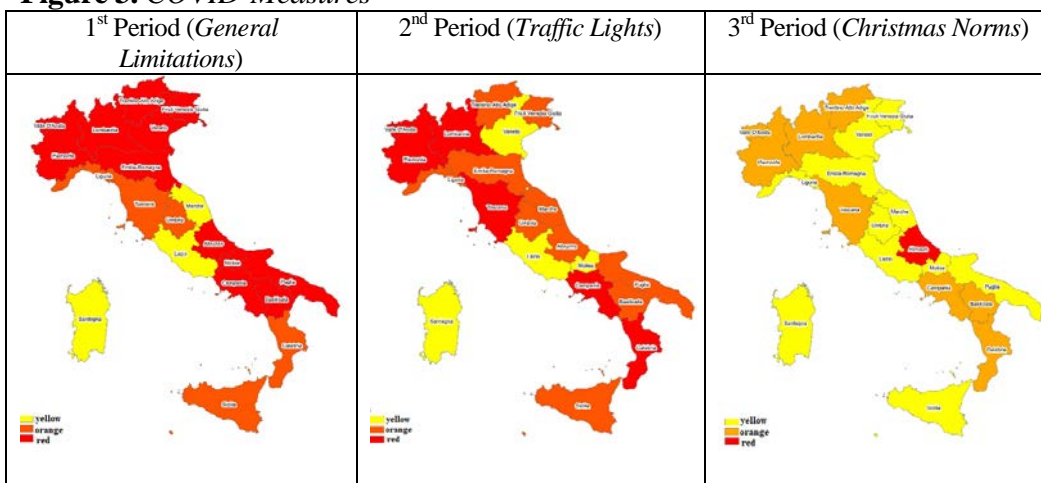
The COVID-Issues Geography

The creation of the *COVID-Issues* involved the analysis of the textual content of the tweets. The lexicometric techniques of the textual data and the application of automatic techniques of topic modeling and clustering were used, using T-lab software, according to the three normative moments considered. Summarizing the thematic content of tweets was appropriate for the extraction of the most relevant topics. The topics were then ordered along a continuum of emotions highlighting negative, neutral, and positive areas of perception related to the emergency. Ordering the emotional categories allowed us to use the same criteria as in the previous geographies, and this, in terms of comparison was very effective (Figure 4). Comparing the three maps, it is evident that the strongly pessimistic communication is concentrated in the Northern Regions and oriented mainly to social and economic problems and protests related to the return of the lockdown. This pessimistic scenario seems to be disappearing in the North and spreading in the rest of Italy and especially in some areas of the South. Here, in fact, discussions about pandemic governance and lockdown management seem to be taking over.

Figure 4. *COVID-Issues*
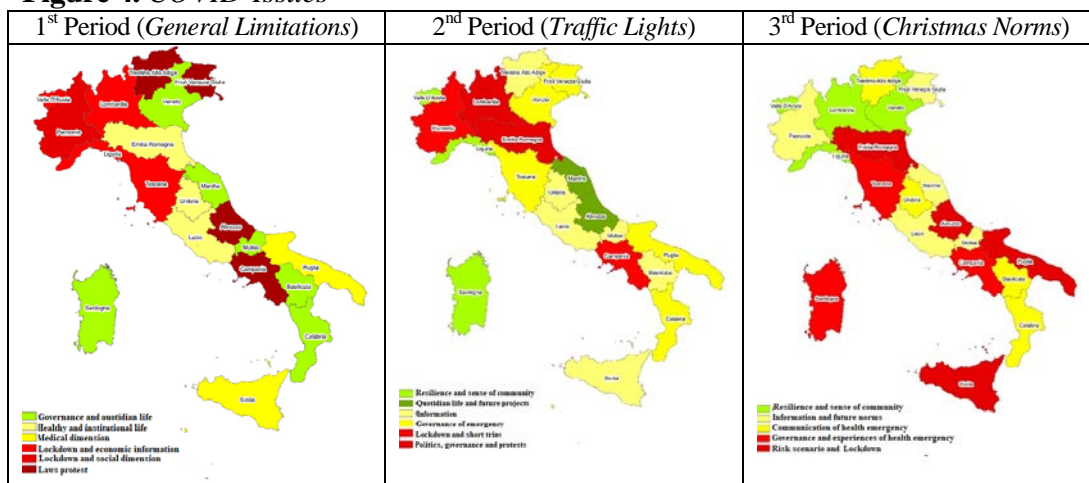
Representing Concordance/Discordance Profiles Starting from Different Geography

The monotonic nature of the three geographies encouraged us to attempt a synthesis to identify the degrees of concordance/discordance between the three levels, for each region. In this sense, the concordance between levels also represents the "expected" situation. For example, a critical outbreak situation (high impact) should predict a context of high restrictions (red zone) and, likely, a narrative with negative sentiment. If this reflection represents the obvious or the expected, a visual plane tool was used to both visually and simultaneously detect clusters of regions with similar concordance/discordance profiles (Figure 5). What we want to show with this visual model for the analysis of concordance and discordance is that the analyzed geographies do not always overlap and that some

situations are recursive, and therefore deserve further investigation. If this could perhaps be the subject of a further article, here we will mainly clarify and test this ambition. Therefore, on the visual plane the x-axis delineates the *COVID-Issues*, the y-axis relates to the *COVID-Spread*, and the *COVID-Measures* are represented by the color of the Regions' labels, which guarantees the three-dimensionality of the plan. In summary, the comparison between the three situations revealed that the second wave of the emergency, at least regarding the three components observed, had a non-linear evolution, describing in some cases even rather contrasting situations.

Figure 5. *Cartesian Plan, "General Limitations"*
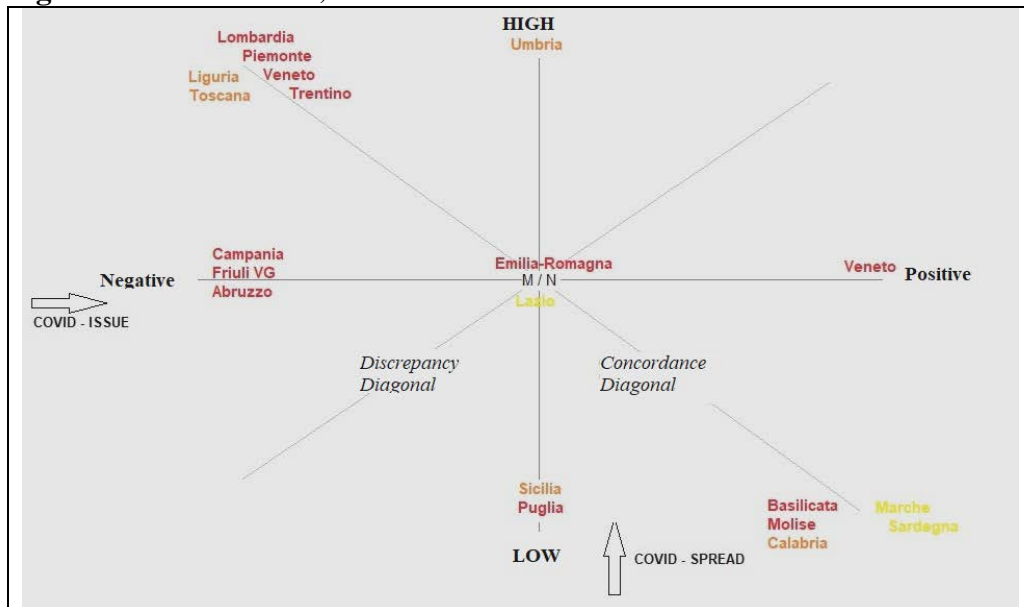
Following an early phase that was not difficult to interpret, with consistent groups of "red" regions with negative narrative polarity and other "yellow" regions with positive polarity (see the figure on the "General Limitations"), two other phases has been followed in which the relationship between contagions, narratives and norms progressively seemed to fragment. Beginning in the "Traffic Light" period, many regions "migrated" from the concordance diagonal to other points in the plane (Figure 6). This is a symptom of how the situation with the passage of time gave way to mixed feelings.

Figure 6. *Cartesian Plan, "Traffic Lights"*
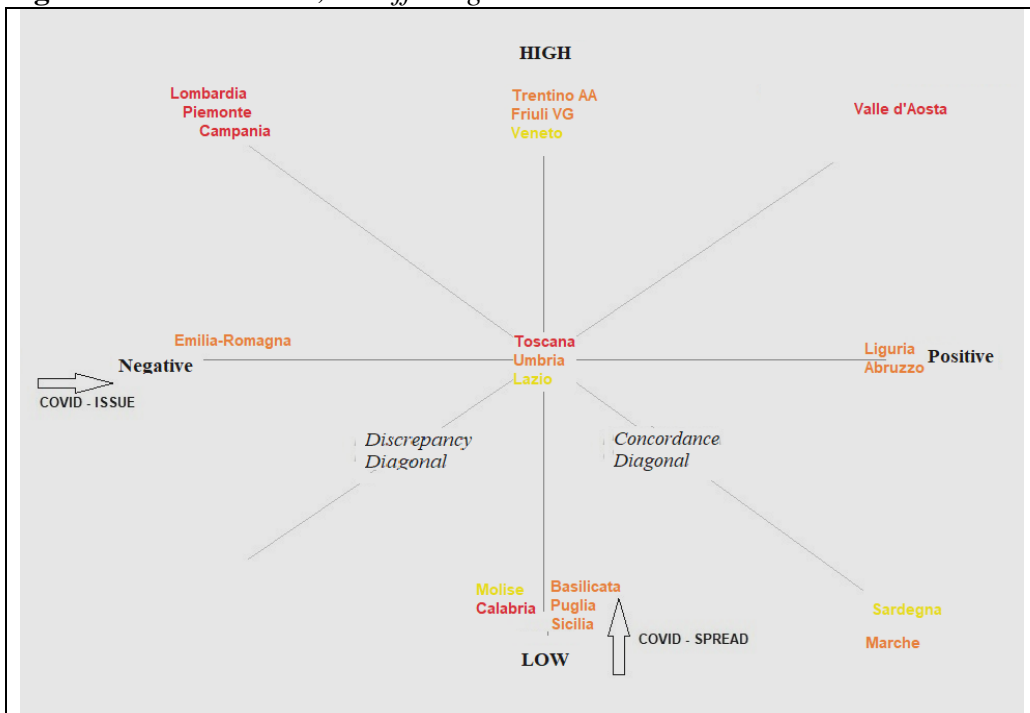
The shift to the last period confirmed that the regions have hardly ever maintained their "positions" along this time continuum of ours, a sign of a rapidly changing situation. Probably the narrative of the emergency has gone in other directions than what the regulatory and epidemiological context could direct (Figure 7).

Figure 7. *Cartesian Plan, "Christmas Norms"*
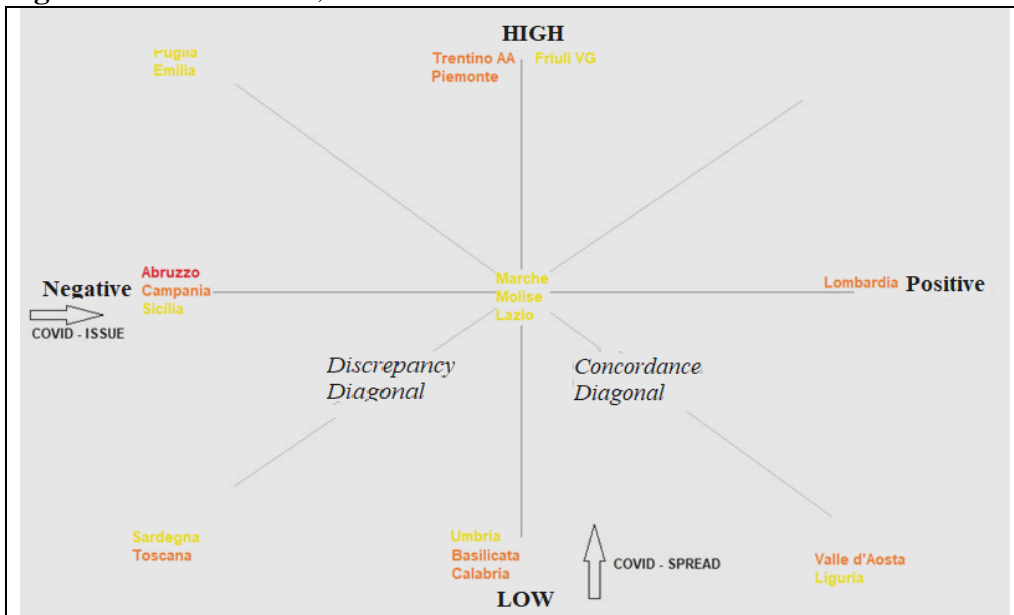
In our case, the spatial analysis together with the analysis of the pandemic perceptions allowed us to hold together two new levels, which in "off-line" research conditions would probably have required a survey with its relative disadvantages (and advantages). Despite the enormous potential we have seen, this new way of doing research with digital data is not without its problems. It is also worth reflecting on the dilemmas that emerge.

Spatial Analysis and Social Media: The Dilemmas of a New Way of Doing Research

Main Conclusion Remarks

The analysis of the geolocalized tweets extracted during the second phase of the pandemic and explored in this study aimed to check if the digital area showed a more cohesive and unified narrative on negative feelings towards the return of the emergency with the change in the trend of its diffusion, which from being concentrated mainly in the North of the country in the second wave linked to the pandemic saw an increasingly homogeneous and fast growing trend.

The analysis carried out returned a complex and articulated picture, formed by multiple perspectives, here traced back to the *COVID-Issues*, *COVID-Spread* and *COVID-Measures* as attributes traced back to the space of their territorial distribution that characterizes them. In fact, this is one of the strengths of social media analysis, the possibility of mapping, defining, and describing all the meanings and associations attributed to a given topic under study.

Are These Limitations in the Use of Geo-Social Media Data or General Features of Analytics on Digital?

Beyond this enormous potentiality offered by the geo-referenced characterization of data deducted from social media with all their characterizations, there are many limitations that can be brought to the attention of the reader. Although for a certain percentage of the data collected on the Twitter platform it is possible to find the data related to the geolocation of the user who shares a specific content, it is not equally obvious that to that type of data the socio-economic-demographic characteristic of the user can also be traced.

More than a limitation of the use of geo-referenced data in the analysis of content coming from social media, this issue is a feature that characterizes all the analyses conducted on the digital scenario which, adopting the perspective that Rogers (2009) declines in the *Digital Methods* approach, are precisely identifiable in the locution post-demographic studies. These are analyses that do not take into account the traditional demographic characteristics of social research such as age, gender, level of education, and so on. In these studies, the individual user is not the unit of analysis of digital research, instead it will be considered as a part of social aggregates that cannot be traced back to socio-demographic categories but to activities (for example, users who comment on a certain Facebook page). What digital research allows us to observe, especially if connected to very specific

criteria such as geolocation, are the activities (writing a post, putting a like, and so on) produced by the interaction between users and digital devices, activities of which the geo-referenced feature remains a simple attribute. The researcher cannot consider digital environments as a window on individual behaviors or personal characteristics of individuals that occur in the physical world, but rather as a strategic observation point of the actors' communication activities that take shape in the digital scenario. Investigating the interactions between users, and between users and devices allows us to capture opinions, value systems, symbols, and identities, that is, the cultural processes that emerge within digital environments. These processes, if usefully located in space can also be represented and mapped in the physicality they take on outside the Net.

What Potential Does This Data Push Beyond?

By immersing ourselves in digital data we can capture the shared culture and perceptions of users with respect to different social phenomena, which in the case of the pandemic can also be connected to meta-attributes that connote the physical space, such as the spread of the virus or the narrowness of the measures introduced to contain the spread of the virus.

The point of contact between the phenomena on the Net and the phenomena outside the Net can be traced back to the reduction of these attributes and meta-attributes to geolocalizable data and representable in the same space that makes the physical territory and the digital space two layers perfectly integrable in the study of complex phenomena. What is possible to do thanks to the use of geolocalized tweets is to analyze the influence of territorial characteristics on the phenomenon analyzed, which in our study is represented by the possible relationship, at the regional level, between the spread of the virus and the type of prevailing narrative.

A first result that this possibility offered by georeferenced social data gave us, led to the evidence that a discordance between the sentiment of the prevailing narratives and the spread of the virus was rarely observed. In regions where daily figures on the virus were high or otherwise concerning, predominantly positive narratives were rarely found and vice versa. In order to identify the second result of the analysis, it is necessary to consider not only the spread of the virus but also the differentiated measures that affected the regions. While it emerges that along the three periods the number of regions in which the linear relationship between virus spread and prevailing narrative type exists decreases, it is also possible to note that in the three periods the "red" regions are rarely characterized by a positive narrative. What emerges is that in the second wave, more than the territorial spread of the virus, it was the type of limitations imposed and, therefore, the risk range that affected regional narratives. This result helps us understand the potential of the transposition into spatial attributes of the characteristics investigated in the study, a transposition that made it possible to concretely answer the proposed research question.

Are other Pitfalls on the Shadow or are There Sediments of Innovation?

The issues related to the post-demographic research and to the transposition of certain given data in meta-spatial attributes are not the only challenge in working with geolocated data coming from social media. In fact, the issue hides other pitfalls relating to the APIs environment and data extraction procedures. APIs (Application Programming Interfaces) are a set of procedures that interface with an application to perform a specific task (extracting Twitter posts, for example). Tweets are public, so there are no privacy constraints. However, we are aware of the risks of automated extraction and the uncritical approach to building large databases (Hernandez-Suarez 2018, Leetaru 2019).

Four limitations can be recalled understanding how challenging it is to work with these procedures and data types.

The first concerns completeness: the APIs work by relevance and not by completeness, so at daily extractions some tweets are often missing from the roll call.

The second concerns the timing or the limit imposed by the proprietary platform of not being able to go back more than 7 days from the date of extraction, an issue solved here by adopting a strategy of extraction in real time. However, this component leads to visible limits.

The third concerns the daily number of extractions that, with the procedure used, is fixed at 18,000 per day. Although it might seem a lot, when you follow the hashtags used as those used for the extraction of the corpus in this study, there is partial amount of data existing in the network.

The last one concerns the limits for each call for which it is possible to extract a maximum of 100,000 tweets per object/hashtag followed. For these last questions, it is possible to find a solution subscribing to packages of extraction. These packages, however, are generally very expensive and are not always explicit in all the characteristics and the limitations that might incur. On the other hand, the Rtweet package add the possibility of selecting the extraction of only geo-referenced tweets, which then limits the number of tweets really useful for analysis and allows us to reach exactly the base of our interest. This extraction is done at zero cost, and this is no small advantage considering that social research is often lacking in terms of research funding.

These dynamics lead to reflection on issues related to how algorithms work and the ability to enter the dynamics of their construction. One of the limitations encountered by analysis with data from social media, in general, remains related to this issue.

All these reflections lead to questions in terms of the representativeness of the achieved results. As noted, both due to the percentage of Italian Twitter users and the extraction limits of the R package, the results are not generalizable to the entire Italian population. The population on the social then has well-defined characteristics that embrace only the most cultured, young, digitally skilled, and particularly interested in the dissemination and sharing of information not necessarily of a personal nature but rather aimed at a broader communication to

large audiences and diverse interests. It is precisely because of this characterization that the opinions extracted from this social often display very polarized dimensions.

However, wanting to precisely explore these oppositions in the *COVID-Issues* putting them in interconnection with the attributes of *COVID-Spread* and *COVID-Measure*, in this study more than assuming the character of limitation this characterization becomes a real potential for georeferenced analysis interpretable as real sediments of innovation in social research on spatial category and diffusive phenomena.

There is no doubt that these are only open questions and far from a precise definition, but they open the field to appropriate debates in social research that is facing the digital component moving in the physical space of the our world where the research is designed and conducted and where problems under study find their space and their dimension.

Acknowledgments

References

Batty M (2013) Big data, smart cities and city planning. *Dialogues in Human Geography* 3(3): 274–279.

Burrows R, Savage M (2014) After the crisis? Big data and the methodological challenges of empirical sociology. *Big Data & Society* 1(1).

Campagna M, Floris R, Massa P (2016) The role of social media geographic information (SMGI) in geodesign. *Journal of Digital Landscape Architecture* 1: 161–168.

Chappell P, Tse M, Zhang M, Moore S (2017) Using GPS geo-tagged social media data and geodemographics to investigate social differences: a Twitter pilot study. *Sociological Research Online* 22(3): 38–56.

Cotesta V, Bontempi M, Nocenzi M (Eds.) (2010) *Simmel e la cultura moderna: la teoria sociologica di Georg Simmel*. (Simmel and modern culture: the sociological theory of Georg Simmel). Volume 1. Morlacchi Editore.

De Falco CC, Punziano G, Trezza D (2021) A mixed content analysis design in the study of the Italian perception of COVID-19 on Twitter. *Athens Journal of Social Sciences* 8(3): 191–210.

Durkheim E (1951) *Suicide: a study in sociology*. New York: Free Press.

Frisby D (2002) *Georg Simmel*. London: Routledge.

Giddens A (1987) *Social theory and modern sociology*. Stanford University Press.

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

Halbwachs M (2018) *La sociologia di Émile Durkheim*. (The sociology of Émile Durkheim). FrancoAngeli.

Halford S, Pope C, Weal M (2013) Digital futures? Sociological challenges and opportunities in the emergent semantic web. *Sociology* 47(1): 173–189.

Hernandez-Suarez A, Sanchez-Perez G, Toscano-Medina K, Martinez-Hernandez V, Sanchez V, Perez-Meana H (2018) A web scraping methodology for by passing twitter API restrictions. In *arXiv preprint arXiv:1803.09875*.

Kitchin R (2014) *The data revolution: Big data, open data, data infrastructures and their consequences*. California: Sage Publications.

Leetaru K (2019, March 7) *How data scientists turned against statistics*. Forbes.

Leszczynski A (2015) Spatial media/tion. *Progress in Human Geography* 39(6): 729–751.

Longley PA, Adnan M, Lansley G (2015) The geotemporal demographics of Twitter usage. *Environment and Planning A* 47(2): 465–484.

Manella G (2009) *L'attualità della scuola ecologica di Chicago*. (The relevance of the Chicago Ecological School). Sociologia Urbana e Rurale.

Mela A (2006) *Sociologia delle città*. (Sociology of the City). Roma: Carocci.

Menduni E (2014) *I media digitali: tecnologie, linguaggi, usi sociali*. (Digital media: technologies, languages, social uses). Gius, Laterza & Figli Spa.

Nurwidyantoro A, Winarko E (2013) Event detection in social media: a survey. In *International Conference on ICT for Smart Society* (pp. 1–5). IEEE.

Rogers R (2009) *The end of the virtual: Digital methods*, Amsterdam University Press.

Rzeszewski M (2018) Geosocial capta in geographical research – A critical analysis. *Cartography and Geographic Information Science* 45(1): 18–30.

Sutko DM, de Souza Silva A (2011) Location-aware mobile media and urban sociability. *New Media & Society* 13(5): 807–823.

Wei H, Zhou H, Sankaranarayanan J, Sengupta S, Samet H (2018) Detecting latest local events from geotagged tweet streams. In *Proceedings of the 26th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems* (pp. 520–523).

Zajczyk F (1991) *La conoscenza sociale del territorio: fonti e qualità dei dati*. (Social knowledge of the territory: sources and quality of data). Volume 186. Milano: FrancoAngeli.

Zhao N, Cao G (2017) Quantifying and visualizing language diversity of Hong Kong using Twitter. *Environment and Planning A: Economy and Space* 49(12): 2698–2701.