A Mixed Content Analysis Design in the Study of the Italian Perception of Covid-19 on Twitter

By Ciro Clemente De Falco[*], Gabriella Punziano[†] & Domenico Trezza[‡]

The digital era and the boom of social, user-generated and freely available and usable content on the Net has brought to the fore a classic technique, accused too often of being highly subjective and requiring a large amount of intellectual work. This technique is Content Analysis, which has seen an unprecedented explosion in recent years. In addition to the incessant flow, speed of diffusion and high volume of today's big data, the attention of social researchers – as well as of anyone interested in drawing information from this enormous proliferation of data – is shifting towards new possibilities. Among these we find that of having a notion of the contents conveyed, of the feelings expressed, of the polarities of big data, but also the chance to extract other information that indirectly speaks of the tastes, opinions, beliefs and transformations behind the behavior of the users of the Net. In fact, secondary data available on the Net, collectable through sophisticated query systems with API or with web scraping software, make it possible to accumulate huge amounts of this dense social data, from which it is possible to try to extract not only trends but real knowledge, in a quantitative as well as in a qualitative manner. This enriches the value of the results that can be produced with Content Analysis and limits, until disappearing, all the critical horizons that have classically left this technique in the shadows, allowing it to find new applicative dignity, validity and reliability (Hamad et al. 2016). In order to explain this evidence, the contribution that we will present attempts to prove that the return of Content Analysis techniques is not only due to the change in the scenario and in the data analyzed, but also to the ability of this technique to innovate and evolve, leading to open analytical perspectives beyond contingent changes. This can be demonstrated through the application of digital mixed content analysis to the recent Covid-19 outbreak and its development of the perception of the Italian population on a specific digital social platform, Twitter.

Keywords: *Digital Mixed Content Analysis Model, digital platform social data, Twitter, Italy, coronavirus.*

Introduction

Complex social phenomena that transit on the Net can be investigated with a technique that has found a renewed place on the social research scene just as big

─────────────────────────────

[*]Research Fellow, University of Naples Federico II, Italy.
[†]Assistant Professor, University of Naples Federico II, Italy.
[‡]Research Fellow, University of Naples Federico II, Italy.

data is making its weight felt: Content Analysis. These phenomena require an epistemological and ontological translation into a multi-comprehensive approach like the Mixed Methods one. This means fitting into the debate introduced by Hesse-Biber and Johnson (2013), for whom "The exponential growth of ''big data'', arising from newly emergent user-generated and streaming digital data from networking sites such as Twitter and Facebook, will place pressures on MM researchers to transform traditional modes of collecting and analyzing data generated from these sites. […] In the coming years, big data methods and analytics may also drive and challenge MM researchers to rethink and innovate and produce new paradigmatic perspectives and research designs and structures. In turn, MM perspectives and praxis can provide models for interpreting and deriving critical insights that that may give a more complex understanding of big data that can bring a set of new questions and understanding to the trending data currently extracted from user-generated social networking sites" (2013: 107).

This is the reason why new applications, new software and new algorithms are being developed, allowing the extraction of the knowledge nested into digital data. All the characteristics of Content Analysis in its qualitative (Schreier 2012) and quantitative (from its birth, Berelson 1952, to the present day, Riff et al. 2019) versions, the contaminations with text mining techniques and the continuous interconnections with network analysis or geographical techniques, are being recovered. This brings to the attention of the social researcher the continuous evolution of the cognitive horizon which allows access to this new digital frontier of Content Analysis, a frontier that has led to the breaking down of the boundaries between qualitative and quantitative approaches, as well as among different disciplines, leading to the birth of forced hybridizations.

It was precisely from these considerations that, given the emergency generated by the spread of Covid-19, with this study we wanted to focus on social data in order to investigate the online perception of one of the populations most seriously affected by this catastrophe: the Italians. Furthermore, we will apply an innovative model devoted to investigating the multivariate nature of social data: a mixed content analysis model born from the reflections in this paper.

The structure of the essay provides the first two paragraphs dedicated to literature review which describes the evolution of content analysis, particularly in relation to mixed methods and the mixed approach in the digital content analysis. The third deals with the methodology, illustrating the analysis techniques and the criteria for the construction of the dataset. The fourth presents the case study of COVID-19 pandemic disease in Italy. The fifth and sixth paragraphs concern the results: the fifth based on a combination of a Lexical Correspondence Analysis (LCA) and a Cluster Analysis (CA) about Covid-19 Italian's perception on Twitter, and the sixth relates to qualitative in-depth analysis of topic and social narratives. The paper ends with a paragraph discussing the results.

Literature Review

Content Analysis: Developments and New Scenarios

Previously used essentially for military purposes, content analysis assumed the status of a research tool in the 1950s after the publication of fundamental texts such as those by Lasswell (1949) and Berelson (1952). Content analysis has 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, Krippendorff 2018, Weber 1990). According to Krippendorf (2018: 13) content analysis is a research technique for making replicable and valid inferences from texts to the contexts of their use. Content Analysis has enabled researchers to sift through large volumes of data with relative ease in a systematic fashion (Stemler 2000). At the same time, the need to face the challenges posed by "old and new" kinds of data retrievable from the web has prompted those who move within the approach to borrow analysis techniques from other disciplines. Therefore, traditional techniques are being accompanied by non-traditional techniques (Herring, 2009). In this regard, two main families can be distinguished in the so-called *web content analysis* (Herring 2009): *digitized methods and digital methods* (Rogers 2013).

Digital methods play a fundamental role in interpreting the evolution of Content Analysis. In general, digital methods can be considered as a set of research and strategy approaches using data produced in digital environments to study socio-cultural changes (Rogers 2009, Caliandro and Gandini 2016). These differ from *virtual methods* (Hine 2005), also known as *digitized methods* (Rogers 2009), paradigms that studying *reality* by adapting social research tools to the Web (for example, the online survey). Rogers (2009, 2013, 2015) was the first author discussing the structure of digital methods. According to Rogers, using digital methods presupposes epistemological choices. This implies knowledge about Internet and the context of the Web network not from an ontological point of view (an entity separate from reality, therefore an object of study) but as a method resource to study people's behavior and social groups. The potentiality of this digital approach to content analysis does not exhaust its potential only in this paradigmatic shift. In fact, it is in the practice of analysis that many other possibilities open up; one is that of the possibility of fruitfully approaching integrated analysis models typical of Mixed Methods Research.

The Mixed Approach

According to Cipriani et al. (2013), talking about the possibility of using Mixed Methods means referring to the "possibility of adapting and coordinating between them more investigation techniques, more types of elementary information, or different paradigms or approaches of a theoretical or methodological nature" (2013, 272). In other word Mixed Methods research centers around researchers being able to collect multiple data using different strategies, approaches and methods.

The desired results of this mixture have the characteristic of being more than the simple combining of the single methods in order to generate grander and more integrated research outcomes (Orina et al. 2015). Many fields of research, with their characterizing methods and techniques, have already experienced the potentiality of the combination of qualitative and qualitative research approaches to pursue the guiding methodological principle of integration. Nowadays, it is not only a question of methodological principle that addresses social researchers, but also the ever-growing relevance of the kind of data used, the information contained therein, the possible multilayers of reality which they lead to, and the undeniable need for integration between these pieces of reality to build ever more complete paths of knowledge. It should also be noted that the crossing of the quantitative-qualitative dichotomy is directly and indirectly supported by perspectives such as those of "live sociology" (Back and Puwar 2012) and of "punk sociology" (Beer 2014). They try to imagine, and direct to at the same time, the development of sociology in the digital world through new, even heterodox forms, compared to consolidated approaches. Furthermore, in a phase in which epistemologically naive approaches (i.e. data-driven) are being asserted, it is important for researchers to affirm their role by emphasizing the importance of facing a cognitive problem through complex approaches capable of giving better answers or to put it to better understand a situation (Creswell 1999).

The Mixed Approach in Digital Content Analysis

Using content analysis in the digital era in order to analyze digital content, such as that on social media, means being faced with old and new challenges. In the current research process, digital content analysis researchers must: formulate their cognitive questions and make the purposes of their analysis explicit; identify the source of the data and contents that they want to analyze; and then select them consistently to the delineated path. The analysis procedures, quantitative or qualitative or both, that they decide to adopt will depend on the hegemony of the research question (mixed methods perspective), but above all on the hegemony of the medium that conveys the contents taken into analysis (digital methods perspective).

Regardless of these considerations, the content analysis process will consist of the coding of raw data according to a classification framework. This framework, on the one hand, will, from the quantitative point of view, claim to extend and generalize the results. From a qualitative point of view, on the other hand, it will attempt to analyze the considered content more in depth. However, thinking that a cognitive question on complex data such as the digital platform social data can involve only one of these sides becomes an understatement. The Mixed Methods perspective is not only necessary, but in a certain sense mandatory.

In this regard, it is sufficient to think that already Holsti (1969), as well as the more recently retrieved claims by Schreier (2012) or Krippendorff (2018), stated that qualitative and quantitative content analysis are not discrete classifications, but rather fall along a continuum, a notion also used by Teddlie and Tashakkori (2011) to define the new horizon for social research methods in the light of the third

approach, the mixed one. Stressing the approach along this continuum allows researchers to extract greater opportunities to gain insight into the meaning of data. Bryman (2012), on this possibility of moving back and forth in the approach, states that, by definition, "content analysis is a research approach that can be situated at the intersection of quantitative and qualitative methods, a place where both methods can meet and that quantifies and qualifies the manifest and latent meanings of the data" (Hamad 2016).

Combining this understanding of content analysis with a solid mixed-methods design could allow the researchers to reach the maximum result from the massive growth of digital texts and multimedia data. Of course, it is true that for researchers using data from social media platforms (e.g., Facebook, Twitter, LinkedIn or similar) there are few guidelines for the collection, analysis, and evaluation of the various types of data.

Methodology

The cognitive interest that moves this study, in addition to demonstrating the return of Content Analysis in the digital environment, can be summarized in three specific research questions:

1. How has the spread of coronavirus directed, polarized and constructed the perception of the phenomenon faced by the Italian users of Twitter?
2. Which actors have been having the most pervasive communication impact on social perception?
3. What is the reasoning that built the social narrative of coronavirus on this social network? With the aim of finding adequate answers to questions so closely related to each other, as anticipated, a mixed content analysis design is required.

The research design at the basis of this proposal can be identified in the sequential nested model by Creswell and Plano Clark (2017). This model, which combines data collection and analysis of a secondary set of qualitative data in a traditional quantitative research design, has the main objective of strengthening the results obtained by integrating them downstream into the process.

It consists of a first quantitative extension phase with the application of an analysis on latency, or the Lexical Correspondences Analysis (LCA) aimed at pulling out of the original set of data the semantic dimensions of synthesis that can, at a later stage, lead to the application of a Cluster Analysis (CA) with T-Lab software aimed at identifying perception profiles of social users on the risk of coronavirus infection. Finally, a qualitative follow-up will help us to develop these results by building a concept map of actors, thematic areas, communication dimensions, and social narratives on the Covid-19 Italian's perception.

The results of the first phase were used to extract the axes or latent dimensions by LCA as the basis for a typology within which the groups obtained with the cluster analysis are projected as useful attributes for delineating the

different emerging profiles. This new distribution of the emerging perception was also enriched by the kind of actors involved and their importance by their number of followers, and the level of sharing and engagement generated by the analyzed materials.

This technique also allowed us to extract the most characterizing set of tweets for each group or cluster retraced with the CA, an extraction that was used to implement a second in-depth qualitative phase of analysis within which we applied a thematic analysis focused on the hermeneutic interpretation of each set of tweets by theme in order to detect new information about the way in which the main differences in communication can be distinguished, as well as kinds and styles of communication, polarity, intensity and direction of the traced perceptions. For each profile deemed relevant, 100 more significant tweets[1] (with in-group high value), were extracted and an in-depth treatment was started on them, which provided for the classification of the contents with the help of NVivo software and the creation of new attributes to be projected in the classification framework that gradually took form with the integrated results of the different quantitative and qualitative phases.

Furthermore, with NVIVO it was possible to reconstruct the maps of the emerging perception controlled on the basis of the arguments and the relationships that were generated among all these elements on the groups brought to the attention by quantitative analysis and this made it possible to also add a relational component, useful for understanding future developments and trends, to the produced framework.

In regards to dataset building, the hashtag extraction was supported by R extract tweet packages (rtweet) to locate current trends in digital content analysis on one of the most popular social media networks, Twitter, which made use of API to collect data. The data collection involved all the tweets about Covid-19 in Italian. It covered the period from March 5-15, when several important decisions relating to Covid-19 mitigation were made (DPCM 4 March 2020). Given the extension of the corpus and the limits relating to the API's Twitter (max 18,000 tweets per day), several daily extractions were carried out. The extraction keys were based on six hashtags, i.e. those that were potential or effective topic trends for the period in question:

- #coronavirusitalia and #coronavirus identify the main theme and, it is assumed, index a more popular and generalist communication on the theme (we could define it as knowledge-oriented);
- #iorestoacasa, #fermiamoloinsieme and #italiazonaprotetta could aggregate communication that was more interested in problem solving, i.e. about measures to reduce the virus risk (so this hashtag group we could call problem solving – oriented).

---

[1]Significant compared to the groups emerging from the CA

The final corpus consisted of about two millions tweets (including retweets). To facilitate mixed design, we decided to work on a more limited sample of 10,000 tweets (without the retweets) randomly extracted respecting the hashtag proportions related to: Tweet daily number and Hashtag groups (Figure 1).

The daily tweet percentages suggest that from the first day of extraction until March 11 there was a progressive increase in 'Covid' tweets. The most active days were those from 8 to 11 (on average with more than 10% of the daily tweets). The high number of tweets is plausibly connected to the implementation of important lockdown orders in Italy, first in the North and then throughout the country. March 11 (after Italy's lockdown) was in fact the day with the most tweets extracted (just over 13% of the entire body). However, there was a slightly decreasing trend after that date.

Figure 1. *Overview Table on Population, Hashtags and Sample of Tweets*

PROPORTIONAL SAMPLING BY N TWEET DAY AND HASHTAG QUOTES							
POPULATION and QUOTES					SAMPLE		
DAY	n tweets	%	#coronavirus #coronavirusitalia	#iorestoacasa #italiazonaprotetta #fermiamoloinsieme	#coronavirus #coronavirusitalia	#iorestoacasa #italiazonaprotetta #fermiamoloinsieme	n tweets
05-mar	61693	2,9%	100,0%	0,0%	288	0	288
06-mar	140002	6,5%	100,0%	0,0%	653	0	653
07-mar	198527	9,3%	99,8%	0,2%	924	2	926
08-mar	234441	10,9%	66,8%	33,2%	730	363	1093
09-mar	281869	13,1%	60,6%	39,4%	796	518	1314
10-mar	262421	12,2%	65,1%	34,9%	797	426	1223
11-mar	284753	13,3%	81,8%	18,2%	1086	242	1327
12-mar	141442	6,6%	69,1%	30,9%	455	204	659
13-mar	143571	6,7%	69,2%	30,8%	463	206	669
14-mar	206125	9,6%	81,7%	18,3%	785	176	961
15-mar	190204	8,9%	85,9%	14,1%	762	125	887
tot.	2145048	100%	77,4%	22,6%	7738	2262	10000

Source: elaboration on R on tweet corpus

As the dataset was building, the automated extraction returned the tweet data related to 88 variables. However, we considered it sufficient to consider just 9 variables, i.e. those consistent with our research design. The 10,000-tweet dataset was built considering: *Display name*, *Verified account*, *Date*, *Time*, *Text*, *Text Width*, *Favorite Count*, *Retweet Count*, *User Followers* and *User Type* (built afterwards).

The *Display Name*, aka Twitter nickname, identifies the individual user and it is useful in defining users' classification.

Verified account is a useful variable for checking official accounts, such as media, opinion leaders, political organizations, etc.

The *Date* and *Time* temporarily place the tweet. The time is useful for specifying the daily range of the tweet, according to the classification: morning, afternoon, evening, night.

The *Text* is returned according to common Twitter standards which has just recently allowed users to exceed the standard 140 characters.

Favorite count, *Retweet Count* and *User Followers* are three quantitative variables discretized by five levels (quintiles). It is plausible to think of the first

two as indicators respectively of the engagement and the sharing levels of the tweet content, while the third variable refers to the popularity of the user and its centrality in the communication arena of the network.

User Type variable was constructed at a later time to define a typology of Twitter user, by a multi-criteria and controlled classification considering the five variables previously seen, i.e. Display Name, Verified Account and quantitative variables (Favorite Count, Retweet Count and User Followers). The variable was coded according to six classes: 'Common User' (lowest level of sharing, follower and engagement value), 'Intermediate User' (second or third level of sharing, follower and engagement value), 'Influencer' (fourth level of sharing, follower and engagement value), 'Top user' (highest level of sharing, follower and engagement value); Political User' and 'Official Information media' (defined only by Display name and Verified Account).

These variables were included in the LCA analysis models as supplementary attributes to better describe the lexical patterns emerging from the textual contents of the tweets.

The Case Background – COVID 19 Pandemic disease

A new coronavirus (COVID-19) was identified in Wuhan, China, in December 2019, declared to be a Public Health Emergency of International Concern on 30 January 2020, and recognized as a pandemic by the World Health Organization on 11 March 2020.

The Italian coronavirus cases surged from hundreds to thousands within two weeks, from a few hundred in the third week of February to over 3,000 in the first week of March, marking the biggest coronavirus outbreak outside Asia (only China and neighboring South Korea had had more cases). The infections in Northern Italy then rose and many other countries in Asia, the Americas, and Europe traced their local cases to Italy.

On March 8, the Italian government announced the lockdown of 11 Italian towns identified as the worst affected, including ten in Lombardy and one in Veneto (DPCM 08 March 2020). Within two days, the quarantine was extended throughout Italy (*iorestoacasa* decree) as COVID-19 cases were detected across the country. The quarantine period would depend upon how soon the number of new cases and deaths would decline. Italy was the first country to announce a nationwide lockdown following the Wuhan coronavirus outbreak.

In such a critical context, models of crisis and emergency risk communication (Beck 2000, Napoli 2007, Reynolds and Seeger 2005, Renn 1992) suggest that it is crucial to understand the perception of risk of the population and the sources of information that they trust to enable effective communication.

Although international and national institutional actors attempted to plan communication strategies for the correct information to mitigate disease, there was a high risk of a spread of fake news, overflow and bad information, especially what was shared on the main social networks (Vaezi and Javanmard 2020). Rumors and misinformation can undermine many public health actions and should be debunked effectively (Betsch et al. 2020).

In our case, the relevant hypothesis is that the spread of information through different institutional or non-institutional sources contributed to polarizing Italian user perceptions about the emergency, from excessive fear and concern to a total lack of interest.

Therefore, it is interesting to construct the main semantic categories of the perception and representation of the disease. In this way, it will also be possible to consider any relationship between the epidemic outbreak and the change in people's perception and feelings to try to improve institutional communication and safety-oriented policies.

Findings/Results

The Quantitative Multidimensional Exploration of the Covid-19 Italian's Perception on Twitter

In this paragraph a multidimensional analysis based on a combination of a Lexical Correspondence Analysis (LCA) and a Cluster Analysis (CA) (Benzecri and Benzecri 1984, Lebart et. al. 1997, Greenacre 1984) was implemented. These are two techniques were used to reduce the space of mining contained in large sets of textual data as well as the dataset that we used for our analysis.

LCA, like all factorial analysis techniques, aims to extract new variables from the original matrix in order to summarize the information it contains. To understand which patterns represent the extracted factors it is necessary to understand which are the modalities of the variables/lemmas enriched by mining these factors in order to identify the concepts that account for the variability that they reproduced. It is for this particular characteristic of the used technique that we were able to extract two new synthetic dimensions of mining that allowed us to interpret the differences among the analysed content. The summary of the results of the LCA was achieved by performing the CA simultaneously, or on the new extracted variables. This technique regroups homogeneous elements within a set of data. In our case, CA served to group tweets characterized by a similar perception expressed with the use of similar words. These perceptions were identified thanks to the mining evidence that emerges from the LCA.

As mentioned above, the first result obtained with the application of the LCA is the delineation of two main synthetic dimensions of mining called factors. These factors can be crossed and used to build a new space of mining generated by this crossing. Figure 2 shows the crossing of these new dimensions, the meaning of which is built in the attraction and repulsion relationships among the active variables used for this analysis (type of user who posts the tweets, day on which they post and time slot and lemmas coming from the tweets text) that we used to describe the synthetized mining found on the new generated factors. Moreover, on the factorial plane obtained, there was also the projection of the cluster that we obtained through the application of a further statistical analysis on this dataset, the CA, which we will describe below.

The first factor is related to the opposition between the *private* and *public sphere* used as *direction of the expressed perception* in the analysed discourses. On the positive semi-axis, we find tweets mainly connected to the individual and private

sphere. Here we have lemmas such as *aperitif, Netflix, home,* and *boring*, that clearly describe individual experience. Meanwhile, on the semi-negative axis we find the terms *health, Companies*, and *OMS*, which refer to the public sphere. The location of user types is decisive. The common user addresses the private sphere while all other users and, in particular political groups or official and administrative bodies, address the public one.

For the second dimension we found an opposition among the focus of the constructed discourses among the tweets. On the semi-negative axis, we found tweets that refer to daily limitations, medical issues, and social measures. Here we have terms such as *responsibility, awareness* and *running away*. On the positive semi-axis, there are tweets related to health service support and communication about the health emergency. The lemmas that we find here are *containment of the Coronavirus, order, Civil protection*, and *measure*. For this particular distribution, the semi-negative axis seems to refer to the many areas affected by the pandemic and therefore to the *social emergency*, those on the semi-positive axis seem to have as their central focus only the health *emergency*. Health and social therefore are the semantic poles of the second factor which is related to the *type of emergency*.

Figure 2. *Factorial Plan on LCA with Active Variables, Lemmas and Cluster*
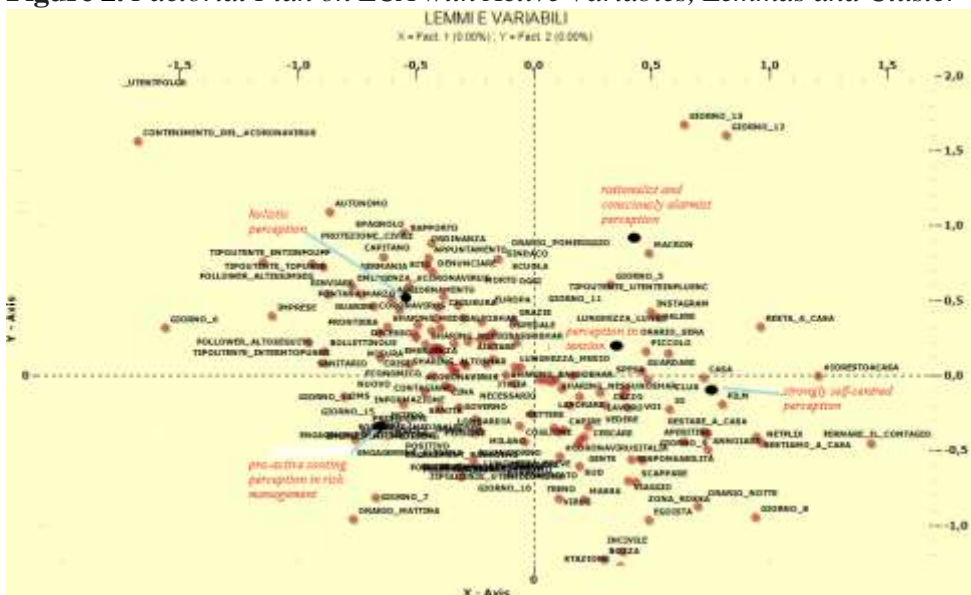
Source: elaboration on T-Lab on our 10,000-tweet sample.

These reflections led us to the question: what were the emerging perceptions regarding the experience induced by the Coronavirus emergency in the analysed corpus? We will try to answer this question by combining the evidence discussed above with the results of the cluster analysis shown in Figure 2. There are five groups extracted from the cluster and each one is characterized by a specific perception of the pandemic that derives from the collectively constructed narration by Twitter users in the first ten days of national lockdown.

The first cluster is located near the centre of the plane, collecting a very high part of the variability of the opinions expressed, but precisely for this reason also more common. It is not a coincidence that the characterizing type of user is the common one

who focuses on very different and seamless aspects. They are willing to describe and understand what is happening (with words like *understand, search, see*) as well as the narration of daily practices (referring to shopping, Instagram reports, etc.). This represents the report of the daily expedients for managing the individual quarantine, and, at the same time they also open to the sense of collective experience for which the motto 'physically distant but close in experience and hopes' holds true, thus also recovering the guidelines of politicians and other great actors who tended to want to give off an aura of relaxation in the general experience. The name that can be attributed to this group is that of *perception in tension* between the most intimate and individual dimension and openness to collective experience.

In the second group which is at the crossroads between a dimension tending to collective-public openness and a propensity towards emphasizing the discourse focused on the health emergency, we find the users to be the local and national political-administrative class, the official information and the top users thus defined for their wide following. The tweets here are the ones with the highest resonance and are mostly centered on a popular narrative. The words relating to this group refer to the multiple aspects of the epidemic crisis: to the actors (such as civil protection, local political actors, institutions, etc.) to the measures (with the use of the words *ordinance, measure, closure*) and to the consequences on the population (such as *deaths, isolation, therapy*). This is a complex narrative that touches various key points of this pandemic precisely because it is the prerogative of the users deemed to be the most influential with afternoon messages that coincided with the circulation of daily update bulletins. It follows that the emerging type can be defined as *holistic perception*.

The third group explicitly refers to the need for support to the healthcare system with words like *support, hospital, and medical staff*. The reconstructed narrative is based on informed opinions about the emergency experienced from a healthcare point of view and a more individual concern weighs not insignificantly. The high information content of these tweets is also motivated by the fact that they are mainly from users believed to be the influencers and therefore able to act on the construction of individual perception starting from the conscious restructuring of the pandemic narration. The result is a *rationalist and consciously alarmist perception*.

The fourth group is the one in which a *strongly self-centred perception* prevails and is in fact moved to the more private and individual side of the first constructed dimension. Here we find the tweets that lead back to the effects on the private sphere of the pandemic. The type of user close to this group is once again the common user who launches a narration focused on everyday things (*Netflix, aperitif*), the experience of quarantine (*boring, new habits, new way of working from home*), the dimension of prayer and recrimination (*awareness, but also running away, selfishness*). These were mostly tweeted in the evening and at night, leaving a glimpse of a search for greater intimacy even in a digital dimension of communication and interpersonal sharing.

The fifth cluster mostly focuses on more general medical emergency issues and technical medical issues. Mainly they were tweeted in the morning as they processed and digested, and condensed the updates released the previous day with the expectations and new ideas for pandemic management in the new day. The result is a *pro-active soothing perception in risk management*.

Furthermore, the division into five groups was functional in paving the way for the development of the qualitative part of this study. For each of the five groups, after identifying the posts that made them up, the most representative 100 posts per group were extracted and, on these, a qualitative analysis with NVivo was conducted on the emerging themes and on the social narratives that we will present below.

The Qualitative In-depth Analysis of Topic and Social Narratives of the Covid-19 Italian's Perception on Twitter

In the previous paragraph we dealt with the reduction of the semantic dimensions contained in the analyzed dataset, in this paragraph, on the other hand, will be dedicated to examining the emerging meanings in these semantic dimensions. This allows us to extract new information about the way in which to distinguish the main differences in the points made by users and emerging themes detectable from the set of analyzed tweets. Along with this, we will look at differences in the building of social narratives that emerge from changes in terms of communication type and style, sentiment polarities, intensity, and direction of the expressed perceptions.

In order to do this, we applied a hermeneutic analysis starting from the classification made possible thanks to the obtained axes or synthetic dimensions of mining generated with the LCA. These dimensions contemplate a first opposition among posts devoted to highlighting the *private* or the *public sphere*, and we also give a connotation as *individual* or *collective horizon* in the perception of the spreading of the pandemic, and a second opposition among the importance assigned to the *social* or *health* dimension of the *emergency*. To reach the profiling of the thematic areas and the type of social narrative traceable among the analyzed short texts, we also considered a series of other dimensions in which it is possible to detect differences or graduations in the way in which these emerge from the texts.

The first kind of differences considered stay in the primary type of communication which gives an impression to the analyzed post especially by highlighting the kind of producer of the message. Most posts could be assigned to an *interpersonal communication* generally conducted by *ordinary people* who give an intimate and emotional connotation to the messages spread (i.e. # day10: *I look out the window and everything seems so unreal. The silence outside reflects the loneliness I live inside* #istayathome- common user). To the opposite side, another considerable number of posts can be attributed to *public and institutional communication* where the main producers are the *institutions*, giving the messages an openness to the collective and the possibility of keeping together the attention focused on very different spheres involved in the pandemic (i.e. #doyourpart *Defend yourself and defend others, wear a mask, keep a distance of one meter* and *limit the outings to those strictly necessary* - institutional user). This openness and dynamic are also attributable to another kind of detected *communication*, the *political* one, used by politicians and local administrators that at the same time sometimes overlap the intimate and emotional connotation of the *interpersonal communication* (i.e. close to everyone's experience #togetherwe willmakeit - political user). The last difference could be traced in the *techno-scientific communication* mainly the prerogative of *scientists, technicians,* and *experts* both in health and in socio-economic measures aimed at curbing the crisis connected to the spread of the pandemic (i.e. *the search for antibodies for a vaccine continues*

#thesearchdoesnotstop #covid-19 - technical user - i.e. *the government is working hard, proposals are being examined to address the socio-economic impacts of this pandemic* - expert user). It follows that these types can be positioned along the continuum between private/individual and public/collective spheres. Therefore, we start from interpersonal communication until we gradually open up to different gradations of collectivity and inclusiveness. Still, along this dimension another continuum is stressed, the one that has the *purposes of the type of communication* as extreme, on the one hand aimed at the *maximum emotional and empathic involvement*, on the other hand aimed at the *maximum rational and conscious involvement*.

Graphically, we could represent that as follow.

**Figure 3.** *Style, Type, and Purposes of Communication*
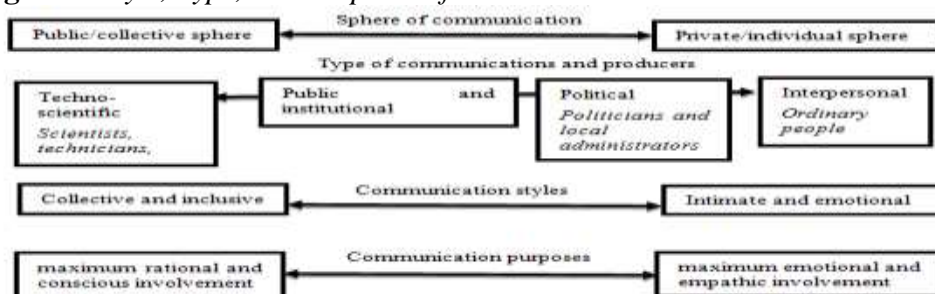
Source: our elaboration.

However, the analyzed posts can also be distinguished on the basis of the *polarity of sentiment expressed*. Although it is possible to identify the extremes of *negative* and *positive*, along this dimension we are not faced with different expressed gradations, but with *different combinations of intensities* in which either polarization is totally canceled, and therefore they are defined as *neutral*, or the polarities combine with each other, we will therefore define them *mixed*. In the text analyzed, if we could assign the neutrality connotation to techno-scientific and public and institutional communications which, on the other hand, are characterized by typical traits of disclosure and information in a constructive and proactive prospective, the mixed connotation is generally assigned to public and institutional communication that share the same traits, intended to be neither alarmist nor optimistic. The extremes of positive and negative are found in the styles of political and interpersonal communication, deliberately more marked and polarized than the other types of communication (i.e. #unitedbutdivided *this pandemic will teach us so much* - common user - *still hundreds of deaths* and #Conte *continues his dictatorship of imprisonment and terror* #businessandpolitics - common user).

Following the generated continua, another one could be produced:

Figure 4. *Sentiment Polarities*
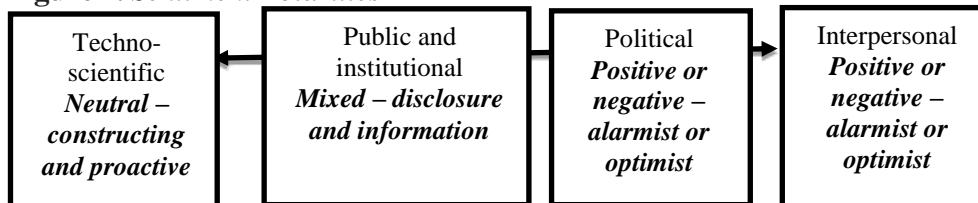
Source: our elaboration.

As far as the *direction of the expressed perception* is concerned, a continuum can be identified in the projection made in the discourses in terms of referring to *past*, *present* or *future*. The reference to the past is more typical of technical-scientific and institutional discourse, aimed at a comparison between what happens in the present and how things have been dealt with and managed in the past (i.e. *will the Ebola vaccine case help in the fight ahead of us today?* - media user). But it is also a typical modality of interpersonal discourse, as it conveys the perception of the present to an anchorage with the past and to that refined return to normality that is typical of the past (i.e. *another friendless day, another empty day* #ridemebacknormal - common user). Experts, institutions and politicians refer to the present to comment on measures and situations, but also ordinary people in concentrating the narratives on how the pandemic is experienced here and now (i.e. *the first effects of the containment measures are starting to show ways out* #everythingwillbefine - media user). On the other hand, if scientists and institutions look with analytical rationality, politicians and ordinary people project hopes and expectations on it (i.e. *the dawn of a new day* #restiamoumani - political user).

The same trend holds the *focus to which the discussion refers* and highlighted as the second dimension of LCA synthesis: the focus on the *social* or *health* dimension of the *emergency*. Whether they are ordinary people, politicians, institutions or experts/scientists, each sphere touched by the emergency is metabolized and returned in the narratives of all the actors involved in different ways and with different intensities. Therefore, unlike the previous ones, these dimensions cannot be stretched along a continuum, but rather belong to the type of topic discussed. And this opens our qualitative analysis to the identification of the thematic areas connected to the characterizations of the discourse and narratives analyzed above.

The main thematic areas that can be traced in techno-scientific communication are: *Public communication on health emergency, Medical issues* and *Informed opinion*. These all belong to the health emergency especially in its impact on the population. The aim is the production and the spread of knowledge among all sectors of society.

The thematic areas most closely connected to the public and institutional communication are: *Institutional and digital communication*, *National measures*, *Measures taken for working, smart working and income*, and *Reflections and comparisons with other countries and risk management plans*. The topics run among social and health emergency concerns. The main aims are seeking answers, reasoning about future impact and activating awareness and responsibility in a population that needs to be better informed and adequately trained.

The political communication thematic areas, on the other hand, are: *Economic and health concerns and hopes*, *Social and political addresses after pandemic*, and *national sentiment*. Also in this kind of communication the topics run among social and health emergency concerns. But this time the main aims are to limit the damage, to active involvement due to the weight of the situation experienced and to build moderate confidence in the future.

Two kinds of thematic areas are more determinant in interpersonal communication. One is more self-centered and the other more collective-oriented. Falling in the first are: *Daily limitations*, *Common sense*, *Losses and dangers*, *New and old habits*,

Quarantine, prayers and recriminations, *Epicenter of the pandemic*. These are more recriminatory, outburst and negative discourses, more passive, characterized by the terror of the unknown, where the citizens are drifting at the mercy of events. Instead, in the second thematic area we find: *National resilience*, *Civic sense and information*, *Health service support*, *Sharing daily things*. These are proactive, support and positive discourses in which it is possible to glimpse a path for the way out. Here the discourse is focused on contingent activities as well as on the future perspectives projected towards returning to normal, focused on the understanding and respect for the rules imposed in a moratorium but proactive way. The dimensions of solidarity and support are determinant.

Before projecting all these characterizations in a general framework of classification suitable for integrating all the results obtained from the quantitative and qualitative phases of analyses, we are now able to synthesize the relationships found among all the recalled dimensions in a concept map.

Figure 5. *Concept Map of Actors, Thematic Areas, Communication Dimensions, and Social Narratives on the Covid-19 Italian's Perception*
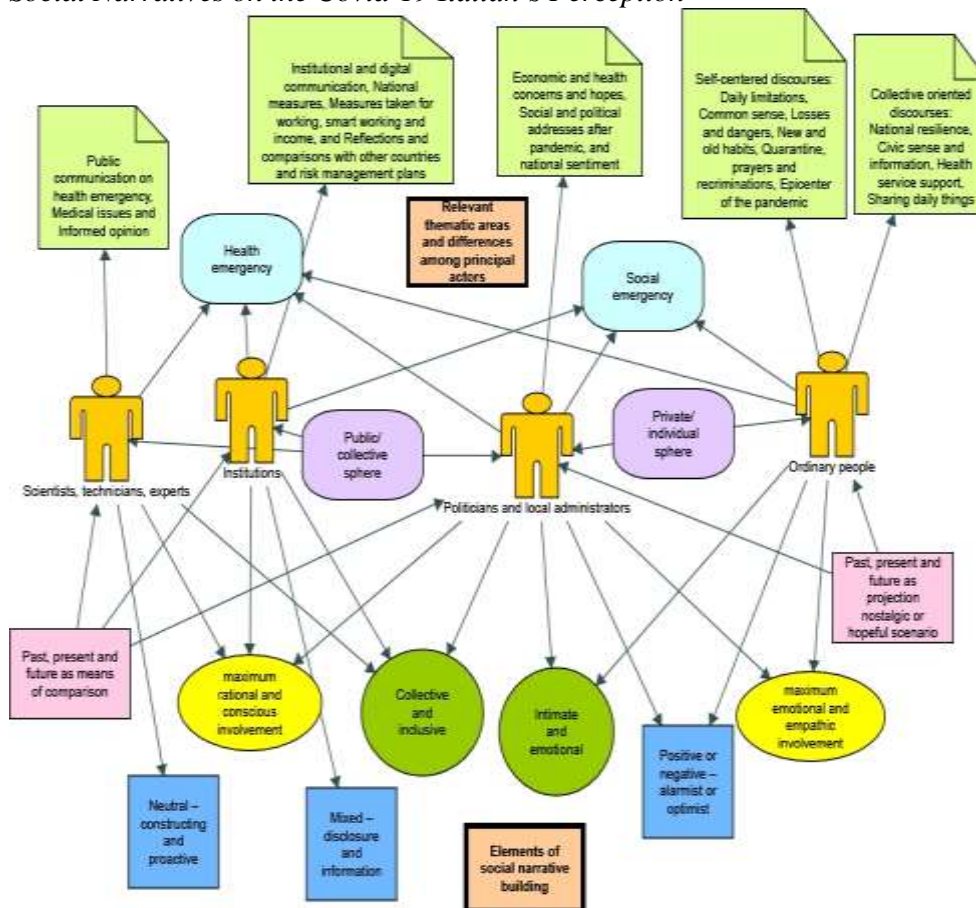
Source: our elaboration with NVivo software.

Discussion and Conclusion

The last step that remains to be done in this analysis involves the integration of the results obtained. With the quantitative procedures, the synthetic dimensions of meaning traced with the application of the LCA have been identified. In order to create a basis for integration, a space of attributes was developed (for example of the conceptual matrices of Calise and Lowi 2010) that crosses these two dimensions, and on this the other elements traced with the other quantitative and qualitative analyses have been projected (see Figure 6).

The horizontal axis shows the contrast between the directions and the projections of the discourse on the *public/collective sphere* on one side and on the *private/ individual sphere* on the other. Instead, the vertical axis represents the opposition in the focus of the speeches, on the one hand on the *social emergency* and on the other on the *health emergency*. In summarizing the terms of the discourse in this way, it is possible to understand which are the *prevailing narratives* for each quadrant obtained using both the groups of perceptions elaborated with the CA and the elements of the construction of the narratives according to the actors who produce them as attributes.

In the upper left quadrant where there is a focus on *health emergency* with prevalent openness to the *public sphere*, the prevalent narrative is the *collective and inclusive narrative*, which emerged during the in-depth analysis of the issues. In this space of meaning two groups of actors with their perceptions can find space. *Politicians* with their proposal for a predominantly *holistic perception*, as well as *ordinary people* when developing their discourse collectively orienting it towards a *corporate perception*.

In the upper right quadrant, which sees the cross between a focus on the *health emergency*, this time addressing a *private and individual sphere*, the prevailing *narrative* is the *rationalist and conscious* one. In this space of meaning we find the *scientists* who propose an *informed perception* and the *institutions* that propose, instead, a *responsible perception*.

In the lower right quadrant born from the cross between a focus on the *social emergency* and discourse oriented towards the *private/individual sphere*, we find a predominantly *intimate and emotional narrative* that is the prerogative of two groups: the *politicians* who propose themselves as representatives of the people, offering an *empathic perception* with each individual, and *ordinary people* who give the most intimate expression of their experience by presenting a self-centred perception instead.

In the lower left quadrant that crosses, once again, a focus on the *social emergency* but this time with openness to the *public and collective sphere*, a *constructivist narrative* prevails. The groups that fall into this are mainly the *scientists* with their speeches focused on a *pro-active perception* in the resolution of the emergency, and the *institutions* that offer reasoning and delineation of future scenarios through a *comparative perception* with other countries, situations and types of emergencies.

As far as research limitations and further developments are concerned, obviously it is necessary to reflect on many points in order to validate the proposed framework, however it is assessable here for its power of theoretical synthesis to restore the vastness of the results in extension and in-depth – qualitative and quantitative –

produced for this study. In particular, we will show this result as a way of integration and visualization of results coming from a *sequential nested mixed content analysis design*, capable of accommodating qualitative and quantitative outcomes and allowing a certain order in the reasoning and interpretation of the – almost always complex – phenomenon chosen as a case study. All this awareness that we are reflecting on refers to a particularly delicate phenomenon whose evolution and impact are ongoing.

To help the reading of such a complex reality, our method proposal can be conceived as a starting point that opens up to new reflections and future developments, continuing to refine the results that can be pursued on both the research paths outlined and on the possibility of their increasingly precise integration. This is because the main research limit lies in the ability to balance the idiosyncrasy of qualitative choices in the pursuit of the extreme objectivity of the qualitative side. Although we tried to manage this feature, it remains a congenital characteristic of the approach to be implemented ontologically, pushing the pragmatic vocation that substantiates the approach and the possibility of presenting a study with the characteristics of the one carried out in these pages.

Figure 6. *Integrated General Model of Classification in the Italians' Perception of Covid-19*
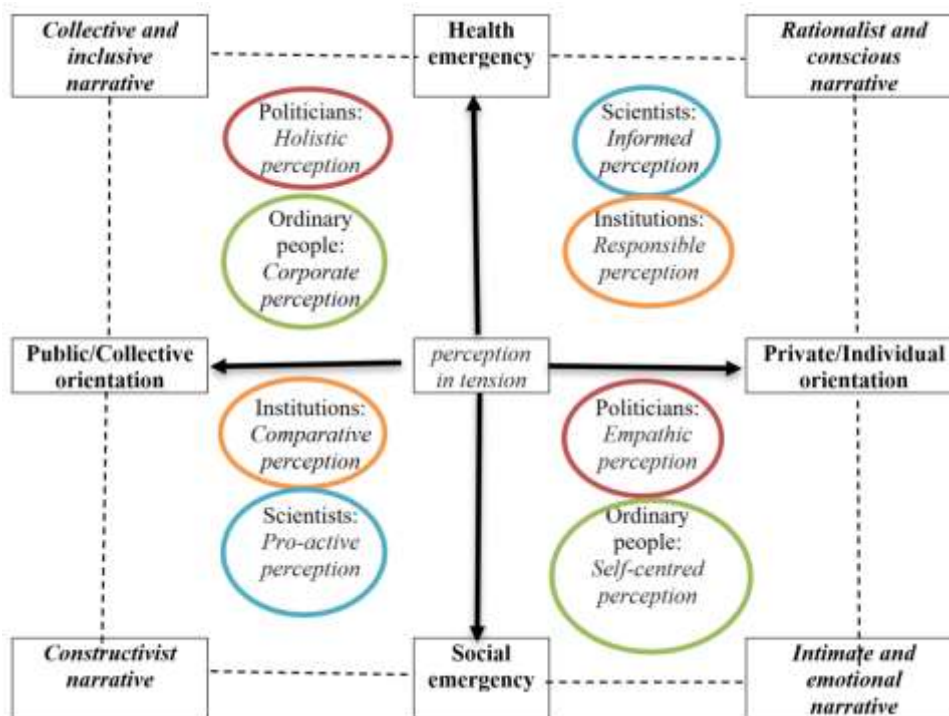
Note: On the axes lie the synthetic dimensions that address social discourses, in the corners the type of narrative, in the quadrants the main actors with their perceptions developed in each specific attribute space.
Source: our elaboration.

Acknowledgments

References

Back L, Puwar N (2012) A manifesto for live methods: provocations and capacities. *The Sociological Review* 60: 6-17. doi=10.1111/j.1467-954X.2012.02114.x

Beck U (2000) *La società del rischio*. Roma: Carocci.

Beer D (2014) *Punk sociology. Houndmills*. UK: Palgrave Macmillan.

Benzécri JP, Benzécri F (1984) *Analyse des Correspondances: exposé élémentaire*. Parigi: Dunod.

Berelson B (1952) *Content analysis in communication research*. Free Press.

Betsch C, Wieler L, Bosnjak M, Ramharter M, Stollorz V, Omer S, ... , Schmid P (2020) COVID-19 Snapshot MOnitoring (COSMO): Monitoring knowledge, risk perceptions, preventive behaviours, and public trust in the current coronavirus outbreak. Retrieved from: PsychArchives. doi= 10.23668/PSYCHARCHIVES.2782.

Bryman A (2012) *Social Research Methods. 4th edition*. New York: Oxford University Press.

Caliandro A, Gandini A (2016) *Qualitative research in digital environments: A research toolkit*. London: Taylor & Francis.

Cipriani R, Cipolla C, Losacco G (2013) *La ricerca qualitativa fra tecniche tradizionali ed e-methods*. Milano: FrancoAngeli.

Creswell JW (1999) Mixed method research: Introduction and application. In *Handbook of educational policy,* Cizek G J (ed). San Diego, CA: Academic Press.

Creswell JW, Clark VLP (2017) *Designing and conducting mixed methods research*. London: Sage publications.

DPCM (DECRETO DEL PRESIDENTE DEL CONSIGLIO DEI MINISTRI) 4 marzo 2020. Ulteriori disposizioni attuative del decreto-legge 23 febbraio 2020, n. 6, recante misure urgenti in materia di contenimento e gestione dell'emergenza epidemiologica da COVID-19, applicabili sull'intero territorio nazionale. (20A01475) (GU Serie Generale n.55 del 04-03-2020).

Greenacre MJ (1984) *Theory and applications of correspondence analysis*. London: Academic Press.

Hamad EO, Savundranayagam MY, Holmes JD, Kinsella EA, Johnson AM (2016) Toward a Mixed-Methods Research Approach to Content Analysis in The Digital Age: The Combined Content-Analysis Model and its Applications to Health Care Twitter Feeds. *J Med Internet Res 2016* 18(3): e60. doi= 10.2196/jmir.5391.

Herring SC (2009) Web content analysis: Expanding the paradigm. In *International handbook of Internet research*, Husinger et. al (eds). Dordrecht: Springer. doi: 10.1007/978-1-4020-9789-8_14.

Hesse-Biber S, Johnson RB (2013) Coming at Things Differently: Future Directions of Possible Engagement with Mixed Methods Research. *Journal of Mixed Methods Research* 7*(2)*:103-109. doi: 10.1177/1558689813483987.

Hine C (2005) *Virtual methods: Issues in social research on the Internet*. Oxford: Berg Publishers.

Holsti OR (1969) *Content Analysis for the Social Sciences and Humanities.* Reading, MA: Addison-Wesley.

Krippendorff K (2018) *Content analysis: An introduction to its methodology*. London: Sage.

Lasswell HD, Leites N, Fadner R, Goldsen JM, Grey A, Janis IL (1949) *The language of politics. Studies in quantitative semantics*. New York: George Stewart.

Lebart L, Salem A, Berry L (1997) *Exploring textual data* (Vol. 4). Dordrecht, Netherlands: Kluwer Academic Publishers.

Napoli L (2007) *La società dopo-moderna: dal rischio all'emergenza*. Perugia: Morlacchi editore.

Orina WA, Mwangi GF, Sitati RN, Nyabola F (2015) Content Analysis and a Critical Review of the Exploratory Design. *Name: General Education Journal 4*(2): 32-45.

Riff D, Lacy S, Fico F, Watson B (2019) *Analyzing media messages: Using quantitative content analysis in research*. New York: Routledge.

Renn O (1992) Risk communication: Towards a rational discourse with the public. *Journal of Hazardous Materials* 29(3): 465-519. doi: 10.1016/0304-3894(92)85047-5.

Reynolds B, Seeger W (2005) Crisis and emergency risk communication as an integrative model. *Journal of health communication* 10(1): 43-55. doi: 10.1080/10810730590904571.

Rogers R (2009) *The end of the virtual: Digital methods* (Vol. 339). Amsterdam: Amsterdam University Press.

Rogers R (2013) *Digital methods*. Cambridge: MIT press.

Rogers R (2015) Digital methods for web research. In *Emerging trends in the social and behavioral sciences: An interdisciplinary, searchable, and linkable resource*, Scott et. al (eds). Wiley online. doi=10.1002/9781118900772.

Schreier M (2012) *Qualitative content analysis in practice*. London: Sage.

Stemler S (2000) An overview of content analysis. *Practical assessment, research, and evaluation 7*(1). doi:10.7275/z6fm-2e34.

Teddlie C, Tashakkori A (2011) Mixed methods research. In *The Sage handbook of qualitative research*, Denzin et al (eds). London: Sage.

Vaezi A, Javanmard SH (2020) Infodemic and risk communication in the era of CoV-19. *Adv Biomed Res* 9:10. doi :10.4103/abr.abr_47_20.

Weber RP (1990) *Basic content analysis* (No. 49). London: Sage.